COMITÉ CONSULTATIF NATIONAL D'ÉTHIQUE
POUR LES SCIENCES DE LA VIE ET DE LA SANTÉ

COMITÉ NATIONAL PILOTE
D'ÉTHIQUE DU NUMÉRIQUE

sous l'égide du
COMITÉ CONSULTATIF NATIONAL D'ÉTHIQUE
POUR LES SCIENCES DE LA VIE ET DE LA SANTÉ

AVIS COMMUN
AVIS 143 CCNE / AVIS 5 CNPEN

PLATEFORMES DE DONNEES DE SANTE : ENJEUX D'ETHIQUE

Avis adopté le 16 février 2023 à l'unanimité des membres présents lors de l'assemblée plénière du CCNE

Avis adopté le 28 février 2023 à l'unanimité des membres présents lors de l'assemblée plénière du CNPEN

Comment citer cet avis :
Plateformes de données de santé : enjeux d'éthique. Avis commun du CCNE et du CNPEN, Avis 143 du CCNE, Avis 5 du CNPEN. Février 2023.

TABLE DES MATIERES

RÉSUMÉ 6

INTRODUCTION 8
1. Contexte de l'autosaisine 8
2. Enjeux de bioéthique et d'éthique du numérique 9

I. DES DONNEES DE SANTE AUX PLATEFORMES DE DONNEES DE SANTE 13
1. Les données de santé à caractère personnel ne sont pas des biens marchands 13
1.1 Principe de l'incessibilité de la donnée de santé à caractère personnel 13
1.2 Différences d'usage et d'origine des données de santé 14
1.3 Données relatives à des spécificités ethniques 15
2. Typologie des infrastructures de données de santé 16
2.1 Bases de données de santé 16
2.2 Entrepôts de données de santé 17
2.3 Plateformes de données de santé 18
2.4 Cohortes 18
2.5 Courtiers en informations (*data brokers*) 18
2.6 Plateformes d'échange de données 18
3. Pourquoi échanger, rassembler et traiter des données de santé massives ? 19
3.1 Motifs de la création de plateformes de données de santé 19
3.1.1 Vers la médecine des 4P 19
3.1.2 Enjeux pour la recherche, l'intérêt public et les organisations privées 19
3.2 Protection de la personne, intérêt public et bien commun 20
3.2.1 Recherche scientifique et intérêt public 20
3.2.2 Marchandisation des données personnelles ou bien commun 21
3.3 Exemples de projets portés par des plateformes de données de santé 22
3.3.1 HYDRO (GIP-PDS) 22
3.3.2 Glucocorticoïdes (GIP-PDS) 22
3.3.3 REXETRIS (GIP-PDS) 23
3.3.4 Pathologies chroniques (Constances) 23
3.3.5 Neuroanatomie du cerveau (UK Biobank) 23
3.4 Trois premières questions éthiques liées à la collecte de données massives 23
4. Principes fondamentaux de constitution des architectures 25
4.1 Sécurité 25
4.2 Interopérabilité 25
4.3 Réversibilité et portabilité 27
4.4 Architectures centralisées ou non 28
4.5 Promotion de solutions techniques ouvertes 29
4.6 Pseudonymisation et anonymisation des données 30
4.7 Bonnes pratiques 31
5. Recommandations 31

II. SOUVERAINETE, AUTONOMIE ET VALORISATION DES PLATEFORMES DE DONNEES DE SANTE 33
1. Enjeux éthiques de la souveraineté des plateformes de données de santé 33
1.1 Une géopolitique des données de santé complexe 33
1.2 Ambivalence de la notion de souveraineté 33
1.3 Une vision libérale et entrepreneuriale pour une souveraineté conquérante 34
1.4 Une vision régulatrice pour une souveraineté protectrice 36
1.5 Une vision européenne fondée sur la notion d'autonomie stratégique 36

1.5.1 L'association Gaia X ... *36*
1.5.2 L'espace européen des données de santé ... *37*
1.6 Tensions éthiques entre visions de la souveraineté et de l'autonomie des PDS .. 37
2. Formes de valorisation des plateformes de données de santé ... 39
2.1 Valorisation sur la base du coût de création et de maintenance ... 40
2.2 Valorisation sur la base des bénéfices futurs espérés ... 40
2.3 Débat européen et français entre les deux formes de valorisation ... 40
2.3.1 Valorisation selon le règlement européen sur la gouvernance des données . 40
2.3.2 Quel financement pour les plateformes de données de santé en France ? ... 41
2.3.3 Le projet de règlement européen sur les données ... 41
2.4 Vigilance concernant les conflits d'intérêts ... 41
3. Recommandations ... 42

III. CONSENTEMENT AU PARTAGE DES DONNEES ET PARTICIPATION CITOYENNE A L'ELABORATION ET A LA GOUVERNANCE DES PLATEFORMES DE DONNEES DE SANTE .. 44
1. Diversité des formes de consentement ... 44
1.1 Consentement libre éclairé et spécifique ... 44
1.2 Autres formes de consentement ... 45
1.3 Données de santé *post-mortem* ... 46
1.4 Enjeux éthiques du consentement ... 47
1.4.1 Avantages des modèles de consentement ... 47
1.4.2 Limites des modèles de consentement ... 48
1.4.3 Tensions et enjeux éthiques ... 48
2. Le choix par défaut *(opt-out)* ... 49
2.1 En France ... 49
2.2 Au Royaume-Uni : « National data opt-out » ... 50
3. Tensions éthiques *opt-in / opt-out* ... 51
4. L'altruisme en matière de données de santé ... 51
4.1 Une nouvelle forme de consentement en vue du bien commun ... 51
4.2 L'altruisme dans le règlement européen sur la gouvernance des données ... 52
4.3 Vigilance sur l'altruisme en matière de données ... 53
5. Pour un écosystème collaboratif pour les plateformes de données de santé ... 55
5.1 Garantir le bien commun ... 55
5.2 Favoriser la participation citoyenne à la gouvernance des plateformes de données de santé via les associations ... 55
5.3 Construire la confiance par l'information, la transparence, la formation et l'accompagnement numérique ... 56
6. Recommandations ... 56

ANNEXES ... 58
Annexe 1 : Recommandations ... 58
Annexe 2 : Membres du groupe de travail ... 61
Annexe 3 : Risques juridiques sur le transfert des données aux États-Unis ... 62
Annexe 4 : Exemples de structures proposant des services de données de santé ... 64
1. SNDS ... 64
2. GIP-PDS ... 65
3. Ouest Data Hub ... 65
4. Entrepôt de l'AP-HP ... 65
5. CASD ... 66
6. Mon espace santé ... 66
7. INCa ... 67
8. Inserm– IReSP - Aviesan ... 67

9. Constances 68
10. UK Biobank 68
11. Dawex 68
12. Salus-Co-op 69
13. Healthbank 69
14. Doctolib 69
Annexe 5 : Données issues de la recherche médicale : quel cadre légal ? 70
Annexe 6 : Auditions 72

RÉSUMÉ

La multiplication des opérations, publiques comme privées, de collecte des données de santé et les complications associées à l'accès à celles-ci a mis en relief leur importance, mais aussi les tensions et les craintes que leurs usages suscitent. Ces données de santé sont de plus en plus souvent rassemblées dans des infrastructures numériques, appelées plateformes de données de santé (PDS), qui proposent en outre des outils d'accès et de traitement.

Le paysage très vaste de ces plateformes privées ou publiques et leur développement croissant dans un cadre aujourd'hui très peu réglementé rendent nécessaire une analyse globale envisageant les conséquences des décisions relatives au recueil, au traitement et à l'utilisation de ces informations sensibles. Par ailleurs, l'architecture matérielle et logicielle ainsi que l'organisation et les ressources humaines dévolues à de telles plateformes méritent d'être interrogées globalement.

Afin d'éclairer les décisions et les politiques publiques relatives à la conception et à la mise en œuvre des PDS, le Comité consultatif national d'éthique (CCNE) et le Comité national pilote d'éthique du numérique (CNPEN) se sont auto-saisis pour mener conjointement une réflexion qui tienne compte des enjeux relevant tant de l'éthique de la santé que de l'éthique du numérique. La réflexion a notamment bénéficié de la collaboration de membres des Espaces de Réflexion Éthique Régionaux (ERER).

Le CCNE et le CNPEN ont construit leur réflexion en s'attachant dans un premier temps à élaborer une définition la plus exhaustive possible de ce que sont les données de santé et à développer, par le biais d'exemples concrets, quels sont leur utilité et leurs usages possibles. Les comités insistent sur le fait que les données de santé ne sont pas des biens marchands mais des attributs des personnes et que par conséquent elles ne peuvent faire l'objet d'un commerce à moins d'être anonymisées, sachant qu'actuellement aucun procédé d'anonymisation n'est certifié. Une typologie des infrastructures est ensuite proposée afin de clarifier le paysage actuel des PDS en dégageant la portée opérationnelle de ces infrastructures et les enjeux éthiques sous-jacents à des choix et des innovations techniques. L'avis s'intéresse ensuite aux enjeux liés à la souveraineté, en attachant une importance particulière à la polysémie de ce terme qui fait se confronter plusieurs visions : libérale et entrepreneuriale, régulatrice et protectrice et enfin une approche alternative dite d'autonomie stratégique. Ces éclairages de la notion de souveraineté permettent de souligner les tensions éthiques soulevées, en s'appuyant sur les principes de bienfaisance, de justice, d'équité des systèmes de soin ou encore d'explicabilité et de transparence. La réflexion se tourne ensuite vers la valorisation des données de santé et identifie deux modèles économiques différents s'y rapportant et soulevant des questions éthiques distinctes.

Enfin, la dernière partie de l'avis est consacrée d'une part aux différents types de consentement à l'usage des données de santé, en particulier à la stratégie par défaut et à l'altruisme en matière de données de santé, et d'autre part, à la participation citoyenne à la gouvernance des PDS. Il apparaît que de nouvelles formes de consentement dynamique sont nécessaires dans la mesure où les données stockées dans les plateformes sont susceptibles d'être utilisées à d'autres fins que celle pour laquelle la personne a initialement donné son consentement. Le CCNE et le CNPEN sont particulièrement sensibles aux questions liées à la participation citoyenne lors de la construction des infrastructures de données de santé puis de leur gouvernance. De nombreuses enquêtes sur ce sujet montrent que la population est peu sensible à ces questions si ces dernières

ne sont pas relayées par des associations de patients, lesquelles jouent un rôle très important dans ce domaine.

Ainsi, au cours de cet avis, le CCNE et le CNPEN font émerger 21 recommandations (dont 3 concernent plus particulièrement la recherche et l'innovation) qui sont regroupées à la fin en fonction des thématiques qu'elles couvrent : la qualité et le partage des données de santé (2 recommandations), l'impact environnemental des PDS (1), leur architecture (4), l'anonymisation des données (1), la souveraineté (4), la valorisation des données (3), et les conditions pour un écosystème collaboratif pour les PDS (6).

INTRODUCTION

1. Contexte de l'autosaisine

En mai 2019, le Comité consultatif national d'éthique pour les sciences de la vie et de la santé (CCNE) soulignait « combien l'accumulation massive de données issues de personnes, comme la capacité accrue qu'a le traitement de ces données de produire de la valeur, nécessit[aient] débat et réflexions éthiques »[1]. La crise engendrée par la pandémie de Covid-19 a mis en relief l'importance de la collecte des données de santé et de l'accès à celles-ci, mais aussi les tensions, les réticences et les craintes que leurs usages suscitent. Ces données de santé sont de plus en plus souvent rassemblées dans des infrastructures numériques, appelées plateformes de données de santé (PDS), qui proposent en outre des outils d'accès et de traitement.

Le développement croissant de plateformes privées ou publiques rassemblant des données de santé – recueillies par des laboratoires, des hôpitaux, des cliniques, des médecins de ville ou des acteurs en marge du parcours de soins – rend nécessaire une analyse globale envisageant les conséquences sur le long terme des décisions relatives au recueil, au partage, à la préservation, au traitement ou à l'utilisation de ces informations sensibles. Cette réflexion n'a pas été close par la création du *Groupement d'Intérêt Public - Plateforme des données de santé* (GIP-PDS) communément appelée *Health Data Hub*[2]. Par ailleurs, l'architecture matérielle et logicielle ainsi que l'organisation et les ressources humaines dévolues à de telles plateformes méritent d'être interrogées globalement.

Ces collections de données personnelles de santé soulèvent des enjeux techniques (de stockage, de sécurité, d'anonymisation/pseudonymisation, de standardisation, de partage, etc.), juridiques (statut de ces données, régime de propriété à adopter, consentement, etc.), mais les questions et les débats qu'elles suscitent, quoique cruciaux, ne doivent pas occulter les enjeux éthiques qu'elles sous-tendent.

Afin d'éclairer les décisions et les politiques publiques relatives à la conception et à la mise en œuvre des PDS, le Comité consultatif national d'éthique (CCNE) et le Comité national pilote d'éthique du numérique (CNPEN) ont souhaité mener conjointement cette réflexion afin de tenir compte des enjeux relevant tant de l'éthique de la santé que de l'éthique du numérique. La réflexion a notamment bénéficié de la collaboration de membres des Espaces de Réflexion Éthique Régionaux (ERER).

Parce que les données de santé représentent un patrimoine immatériel capital, leur disponibilité et leur exploitation numérique ne doivent pas aller à l'encontre des droits fondamentaux des personnes. Un équilibre doit être trouvé entre les exigences portées par l'intérêt public et celles garantissant le respect de la vie privée.

La construction de bases de données de santé, de cohortes, d'entrepôts et de PDS représente un investissement important, souvent financé par la solidarité nationale. Il est important de préciser les valeurs sur lesquelles nous voulons continuer de bâtir notre système de santé. Cela nécessite de définir le statut juridique de ces données et les modes de rétribution des différents acteurs qui ont contribué à la construction, la maintenance et l'utilisation de ces bases, cohortes, entrepôts et PDS. L'articulation et l'harmonisation des règles de sécurisation et d'accès aux informations doivent aussi être anticipées.

---

1 CCNE, avis n°130, 29 mai 2019, *Données massives et santé : une nouvelle approche des enjeux éthiques*, 94 p.

2 Jugement du tribunal administratif de Paris du 22 octobre 2022 concernant l'appellation du GIP-PDS.

La création d'un guichet unique centralisant un très grand nombre d'informations d'une part et la mise en réseaux de multiples plateformes indépendantes et spécialisées d'autre part sont des solutions qui présentent chacune des avantages et inconvénients en termes de gouvernance, de sécurité, de coût financier, de possibilité de valorisation ou encore de capacité de calculs pour le traitement des informations.

Enfin, au-delà du consentement individuel à mettre à disposition des données personnelles de santé des personnes en général, et des patients en particulier, il y a un enjeu particulier de démocratie en santé qui pose les questions de la participation citoyenne à la gouvernance des PDS et du rôle des associations de patients ou d'aidants dans l'élaboration de projets de recherche.

Il convient d'interroger la manière dont ces choix stratégiques peuvent s'inscrire dans le paysage plus global de notre système de santé et de ses valeurs.

Il est important que la France ait une politique claire pour à la fois revendiquer une réelle souveraineté sur ses données de santé et contribuer à l'effort européen et international dans le domaine de la santé publique. Elle doit donc se donner les moyens techniques suffisants pour être en mesure de traiter et d'analyser ces informations aux niveaux national et européen.

Ainsi, elle pourra penser et construire la place de ces nouvelles ressources au sein de son système de santé et organiser le partage de leurs bénéfices potentiels, dans le respect des valeurs de solidarité, de dignité humaine, de justice et d'autonomie dont il est porteur. Une comparaison internationale dans ce domaine peut être utile pour appréhender les atouts, identifier les difficultés à surmonter et les partenariats à susciter.

Encadré n°1 - Données de santé, Bases de données de santé, Plateformes de données de santé

Différentes notions sont utilisées dans cet avis, il nous semble important de caractériser d'emblée notre périmètre de réflexion en définissant leurs contours.

Données de santé : données à caractère personnel relatives à la santé (physique ou mentale) des patients, recueillies par des laboratoires, des hôpitaux, des cliniques, des médecins de ville ou des acteurs en marge du parcours de soins.

Base de données de santé : ensemble structuré et organisé permettant la conservation de grandes quantités d'informations portant sur un domaine spécifique dans le champ de la santé, afin d'en permettre l'exploitation.

Plateforme de données de santé : infrastructures numériques, privées ou publiques, proposant l'accès aux données de santé et leur traitement.

2. Enjeux de bioéthique et d'éthique du numérique

La réflexion conjointe du CCNE et du CNPEN a été nourrie par les valeurs communes et spécifiques de la bioéthique et de l'éthique du numérique. Les réflexions apportées par les différents contributeurs de l'ouvrage *Pour une éthique du numérique*[3] apportent un éclairage sur ce sujet.

[3] Comité national pilote d'éthique du numérique - *Pour une éthique du numérique*. É. Germain, Cl. Kirchner, C. Tessier, PUF 2022, ISBN 978-2-13-083348-2.

Encadré n°2 : Le manifeste du CNPEN « Pour une éthique du numérique »

Le manifeste du Comité national pilote d'éthique du numérique (CNPEN), élaboré à l'occasion de son séminaire annuel (commun au CCNE) les 15 et 16 septembre 2020, et publié en avril 2021, identifie les fondements de la réflexion en éthique du numérique.
Il tient notamment compte du phénomène de systématisation de la quantification et de l'évaluation des activités humaines, qui vient interroger la relation à la connaissance et à la mémoire. L'éthique du numérique est ainsi mise au défi de prendre en compte la remise en jeu de la manière dont nous considérons l'autonomie humaine. Le manifeste soulève également que ces technologies et les modèles économiques qui les portent bouleversent les différents espaces de souveraineté. Ainsi, la réflexion éthique menée par un comité est essentielle aux personnes et institutions qui développent, commercialisent, réglementent et utilisent des technologies numériques.

On peut observer un grand nombre de points communs entre les réflexions menées au CCNE et au CNPEN. La bioéthique et l'éthique du numérique (ou cyberéthique) partagent quelques principes : respect de la dignité humaine et de l'autonomie, de non-malfaisance, d'équité et de justice. Ces deux pratiques de réflexion éthique se recoupent en de nombreux points : elles reposent sur une réflexion collective et pluridisciplinaire, s'attachant à évaluer les innovations scientifiques prises dans leur contexte d'usage ; elles identifient les tensions que ces innovations font surgir entre ces principes et cherchent à évaluer des conséquences que l'on peut raisonnablement leur attribuer.

Ces deux champs de réflexion éthique ont cependant chacun leurs particularités. La plupart des auteurs sont d'accord sur ce point, même s'ils ne pointent pas tous les mêmes différences qui portent d'abord sur les principes qui régissent chaque activité. Ainsi, on peut remarquer que si la bioéthique exige de la pratique médicale qu'elle vise le bien du patient, nul ne demande aux développeurs d'outils numériques de se soumettre à une telle exigence[4]. Ces derniers sont conçus pour répondre, dans le cadre d'une activité à but économique, social ou de gestion publique, à des besoins des usagers qui sont plus ou moins construits. Ceci amène à concevoir l'éthique du numérique comme une première étape vers la construction d'une règlementation, encore trop parcellaire, visant à protéger les usages du numérique, alors que la bioéthique bénéficie déjà d'appuis juridiques bien établis.

Mais, sous un autre angle[5], on peut considérer que la différence entre la bioéthique et la cyberéthique provient du fait que, si les pratiques liées au soin parlent d'abord de l'humain, les outils numériques manipulent d'abord de *l'information*[6] en grande quantité et de nature hétérogène, ce qui soulève des problèmes singuliers. En particulier, le numérique pose des questions de surabondance ou « d'infobésité » qui peuvent conduire à de la désinformation ; il rend possible l'intrusion de la sphère publique dans la sphère privée ou familiale (cas des harcèlements). Cela rend nécessaire une évaluation de la qualité et de la robustesse des outils (qui se posent aussi en médecine mais où elles sont beaucoup mieux cadrées règlementairement) ; et enfin le numérique engendre des questions liées à

[4] R. Chatila - Chapitre « Bioéthique et éthique du numérique : une hybridation paradoxale » in *Pour une éthique du numérique*, *opus cit*.

[5] C. Froidevaux, G. Adda, Chapitre « Regards croisés sur la cyberéthique et la bioéthique », in *Pour une éthique du numérique*, *opus cit.*.

[6] Les professionnels de santé manipulent aussi de l'information, mais elle est collectée avec un objectif défini (améliorer la santé du patient) alors que les professionnels du numérique manipulent de très grandes quantités de données très hétérogènes collectées pour des objectifs divers et réutilisables pour d'autres objectifs dont certains sont liés à la santé.

la transparence des dispositifs, formulées en cyberéthique en terme d'explicabilité (c'est-à-dire que le fonctionnement d'un outil doit pouvoir être présenté de façon compréhensible pour une personne raisonnablement lettrée). La question se pose aussi en bioéthique mais constitue un défi plus criant en éthique du numérique avec l'avènement d'algorithmes d'apprentissage machine dont il est souvent difficile de comprendre le fonctionnement et qui sont parfois qualifiés d'opaques.

Face à ces spécificités de l'éthique du numérique, cet avis se propose d'emprunter une voie maximaliste : il prendra en compte non seulement les principes communs aux deux types de réflexion éthique, mais aussi ceux qui sont spécifiques à l'une ou à l'autre.

Principes d'éthique biomédicale	Principes d'éthique du numérique
• **Principe d'autonomie :** obligation de respecter les capacités de décision et le consentement des personnes autonomes ;	• **Principe d'autonomie** : préserver la capacité humaine d'agir sur les outils et les données ;
• **Principe de bienfaisance :** obligation de procurer des bénéfices et de mesurer les bénéfices par rapport aux risques ;	• *Si le patient est au cœur de l'éthique biomédicale, tous les systèmes numériques ne sont pas conçus pour le bien de leurs utilisateurs.*
• **Principe de non-malfaisance :** obligation d'éviter de nuire ;	• **Principe de non-malfaisance** : ne pas nuire ni exacerber un mal (sûreté, sécurité, robustesse technique) ;
• **Principe de justice :** obligation d'équité, non-discrimination, juste distribution des bénéfices et des risques.	• **Principe de justice** : équité, réduction des biais, non-discrimination, proportionnalité ;
• *Le principe d'explicabilité est présent dans la pratique médicale en lien avec le consentement éclairé.*	• **Principe d'explicabilité** : transparence, interprétabilité, traçabilité, auditabilité. Principe fondamental avec l'avènement de l'apprentissage profond.
D'après l'ouvrage *Pour une éthique du numérique*[7] qui s'appuie d'une part sur les principes de bioéthique de Beauchamp[8], d'autre part sur le rapport du *Groupe d'experts indépendants de haut niveau sur l'intelligence artificielle* constitué par la Commission Européenne en juin 2018[9].	

[7] R. Chatila - Chapitre « Bioéthique et éthique du numérique : une hybridation paradoxale » in *Pour une éthique du numérique*, *opus cit.*, p.34

[8] Beauchamp T.L., Childress J., (1979, 1st edition), *Principles of biomedical ethics*, New York: Oxford University 314 p. ; Beauchamp T.L., (2003), *Methods and principles in biomedical ethics*, J Med Ethics, Oct;29(5):269-74. doi: 10.1136/jme.29.5.269. PMID: 14519835; PMCID: PMC1733784.

[9] Groupe d'experts de haut niveau sur l'intelligence artificielle, Commission européenne, *Lignes directrices en matière d'éthique pour une IA digne de confiance*, avril 2019 - Voir : https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Récemment le rapport du Conseil d'État au Premier ministre sur l'Intelligence artificielle (IA) et l'action publique[10] proposait de retenir sept principes qui correspondent à ceux du tableau ci-dessus en les précisant :

- Primauté humaine sous plusieurs aspects : bénéfice, supervision, non-dépendance et assistance, acceptabilité sociale de « l'erreur machine », forme indirecte de l'erreur humaine ;
- Performance : indicateurs (exactitude, temps de réponse, taux de satisfaction), niveaux acceptables et déterminants de cette performance ;
- Equité et non-discrimination : type d'équité, risque de biais algorithmiques, accessibilité et universalité ;
- Transparence : droit d'accès à la documentation du système, exigence de loyauté, d'explicabilité, conception transparente et auditabilité ;
- Sûreté (cybersécurité) ;
- Soutenabilité environnementale ;
- Autonomie stratégique.

Encadré n°3 : Avis n°130 du CCNE « Données massives et santé : une nouvelle approche des enjeux éthiques »

En réponse à une saisine de la ministre des Affaires sociales et de la Santé à propos des questions éthiques liées aux activités de recueil et de traitement des données massives dans le domaine de la santé, le CCNE publiait en mai 2019 son avis n°130 intitulé « Données massives et santé : une nouvelle approche des enjeux éthiques ». Dans cet avis introduisant les enjeux d'éthique soulevés par la complexité de la révolution numérique, le CCNE dresse un état des lieux des données massives dans le champ de la santé et propose une analyse des questions éthiques qui ont émergé avec le développement du recueil, du traitement et de l'exploitation numérique de données de santé. Le lecteur pourra se reporter à cet avis pour approfondir la question de la diversité des acteurs, des données, de leurs objectifs et l'enjeu de ces profondes mutations en matière de protection des données des personnes.
Nous rappelons ici trois des recommandations de cet avis n° 130 du CCNE qui sont particulièrement pertinentes pour le présent avis :
Recommandation n° 10 : le CCNE préconise le développement de plateformes nationales mutualistes et interconnectées ;
Recommandation n° 11 : le CCNE estime qu'en matière de recherche, l'impératif éthique doit être adapté à chaque situation particulière, de manière à justifier une relation de confiance entre les titulaires des données et ceux qui y ont accès et qui les traitent ;
Recommandation n° 12 : le CCNE considère qu'il est nécessaire de faciliter le partage des données de santé pour les besoins de la recherche.

[10] Conseil d'Etat, *Intelligence artificielle et action publique : construire la confiance, servir la performance*, Etude à la demande du Premier ministre, 31/03/2022, https://www.conseil-etat.fr/publications-colloques/etudes/intelligence-artificielle-et-action-publique-construire-la-confiance-servir-la-performance.

I. DES DONNEES DE SANTE AUX PLATEFORMES DE DONNEES DE SANTE

1. Les données de santé à caractère personnel ne sont pas des biens marchands

1.1 Principe de l'incessibilité de la donnée de santé à caractère personnel

Le Règlement général sur la protection des données (RGPD) du 27 avril 2016, appliqué depuis mai 2018, donne une définition assez large des données de santé (article 4. 15 et considérant 35). Nous reprenons ici la présentation de la Commission nationale de l'informatique et des libertés (CNIL) dans sa note « Qu'est-ce qu'une donnée de santé ? »[11] :

> « Les données à caractère personnel concernant la santé sont les données relatives à la santé physique ou mentale, passée, présente ou future, d'une personne physique (y compris la prestation de services de soins de santé) qui révèlent des informations sur l'état de santé de cette personne. Cela comprend donc les informations relatives à une personne physique collectées lors de son inscription en vue de bénéficier de services de soins de santé ou lors de la prestation de ces services […]) ; les informations obtenues lors du test ou de l'examen d'une partie du corps ou d'une substance corporelle, y compris à partir des données génétiques et d'échantillons biologiques ; les informations concernant une maladie, un handicap, un risque de maladie, les antécédents médicaux, un traitement clinique ou l'état physiologique ou biomédical de la personne concernée (…). Cette définition permet d'englober certaines données de mesure à partir desquelles il est possible de déduire une information sur l'état de santé de la personne ».

La donnée de santé se définit ainsi par sa finalité.

En droit français comme en droit européen, **la donnée à caractère personnel ne se rattache pas au droit des biens mais aux droits de la personnalité**. Le RGPD a encore récemment consacré cette conception inspirée du personnalisme[12]. En revanche, les bases de données sont protégées par le droit d'auteur ou de la propriété intellectuelle.

La donnée à caractère personnel est un élément d'information relevant de la liberté d'expression, ce qui implique qu'elle ne peut ni faire l'objet d'une appropriation ni être cédée. **Qualifiée d'attribut de la personne**, et relevant donc des droits de la personnalité, elle est très étroitement liée à la vie privée[13]. S'agissant tout particulièrement de la donnée personnelle de santé, elle relève de ce qu'il y a de plus intime dans le fonctionnement du corps humain[14] ; à ce titre elle est entourée de plus de garanties que d'autres données, notamment en ce qui concerne son traitement. La personne jouit d'une protection très forte, qui va d'une certaine manière à l'encontre de sa liberté, dès lors qu'il lui est interdit

---

[11] CNIL, « Qu'est-ce qu'une donnée de santé ? » [https://www.cnil.fr/fr/quest-ce-ce-quune-donnee-de-sante].

[12] D'inspiration kantienne (Charles Renouvier) et spiritualiste (Nicolas Berdaiev, Emmanuel Mounier), le personnalisme fait de la personne humaine, être rationnel, volontaire et en relation – par opposition à un individu égocentrique – le centre et le fondement de toute possibilité de connaissance et de créativité et d'engagement dans le monde. Voir https://fr.wikipedia.org/wiki/Personnalisme

[13] Article 9 du Code civil : « Chacun a droit au respect de sa vie privée ».

[14] Article 16-1 du Code civil : « Chacun a droit au respect de son corps, le corps humain est inviolable. Le corps humain, ses éléments et ses produits ne peuvent faire l'objet d'un droit patrimonial ». Article 16-5 : « les conventions ayant pour effet de conférer une valeur patrimoniale au corps humain, à ses éléments ou à ses produits sont nulles ».

de vendre son corps, ou un organe de son corps, ou une information liée à ce corps. C'est ainsi que l'article 1111-8 du Code de la santé publique prohibe, sous peine de sanction pénale, « tout acte de cession à titre onéreux de données de santé identifiantes, directement ou indirectement, y compris avec l'accord de la personne concernée ».

Toutefois, la protection de la donnée à caractère personnel, qui s'est récemment renforcée en droit français puis dans le droit de l'Union européenne, n'est pas uniquement défensive et protectrice de l'individu contre lui-même et les institutions, mais donne aussi à la personne un rôle actif. Ainsi, la loi pour une République numérique du 7 octobre 2016 a introduit dans l'article 1er de la loi du 6 janvier 1978 un second alinéa reconnaissant à toute personne le « **droit de décider et de contrôler** les usages qui sont faits des données à caractère personnel la concernant ». Cette évolution se manifeste notamment par la création de nouveaux droits, comme la possibilité de donner des directives sur le sort des données de la personne après le décès, le droit à l'oubli et le droit « **à la portabilité** des données ».

1.2 Différences d'usage et d'origine des données de santé

La donnée de santé relève d'une personne, mais aussi du professionnel de santé qui a examiné cette personne, et de celui qui interprète la donnée. Ainsi cette donnée de santé est enrichie par d'autres acteurs que la personne à laquelle elle se réfère.

La CNIL, dans sa note sur les données de santé[15], distingue trois catégories de données : (i) les données de santé par nature telles que celles qui portent sur les maladies, (ii) les données qui deviennent des données de santé après croisement avec d'autres données, dans la mesure où elles permettent de tirer une conclusion sur l'état de santé ou le risque pour la santé d'une personne, et (iii) les données dont l'usage sur le plan médical en fait des données de santé. On parle aussi de données de santé pour un usage primaire dans la première catégorie et pour des usages secondaires dans les deuxième et troisième catégories.

Les données de santé sont des données personnelles très spécifiques et sensibles. Dans son considérant 51, le RGPD précise que : « Les données à caractère personnel qui sont, par nature, particulièrement sensibles du point de vue des libertés et des droits fondamentaux méritent une protection spécifique, car le contexte dans lequel elles sont traitées pourrait engendrer des risques importants pour ces libertés et droits ».

En vertu de l'article 9 du RGPD, le traitement des données de santé est en principe interdit sauf si la personne a donné son consentement. Il est également permis même si elle ne l'a pas donné, dans un certain nombre de cas, en particulier la préservation des intérêts vitaux de la personne concernée, si elle se trouve dans l'incapacité de donner son consentement, et les besoins de la gestion des systèmes et services de santé ou de protection sociale, de la médecine du travail, de la médecine préventive, des diagnostics, des soins, et du traitement[16].

On peut noter qu'en France l'utilisation des données de santé en vue de recherches impliquant la personne humaine est soumise à un cycle de validation relativement complexe qui est présenté en annexe (Annexe 5).

À ces données de santé obtenues dans un contexte médical, on peut coupler d'autres données obtenues par des capteurs d'objets connectés tels que des montres connectées

[15] CNIL, « Qu'est-ce qu'une donnée de santé ? » [https://www.cnil.fr/fr/quest-ce-ce-quune-donnee-de-sante]

[16] CNIL, « Recherche médicale : quel est le cadre légal ? », [https://www.cnil.fr/fr/quelles-formalites-pour-les-traitements-de-donnees-de-sante-caractere-personnel]

portées par des personnes, malades ou en bonne santé, qui échappent au diagnostic médical proprement dit mais peuvent fournir une continuité d'enregistrement de la fréquence cardiaque, de l'électrocardiogramme, de la sudation, du niveau de stress, du sommeil etc., dans des conditions de vie très variées : repos, travail, sport. On parle alors d'« applications de bien-être », mais ces données sont souvent pertinentes et complémentaires des données de santé proprement dites.

La possibilité de les faire entrer dans des PDS, et en particulier dans le dossier médical numérique personnalisé, comme dans *Mon espace santé*, est alors tentante mais pose des problèmes spécifiques de droit d'accès pour les personnes concernées et pour les professionnels de santé, et de mise à disposition à des tiers par les plateformes numériques qui les stockent et éventuellement les traitent par des algorithmes pour en tirer des diagnostics personnalisés ou des données statistiques de santé publique. À ce propos *Mon espace santé* a une procédure de référencement des applications mises à disposition des usagers, mais n'opère pas de choix entre des applications concurrentes. Ce choix incombe donc à l'usager qui peut se sentir démuni par rapport à une surabondance d'offres.

Le récent avis du CCNE et du CNPEN sur le diagnostic médical et l'IA[17] spécifie : « Des données plus subjectives peuvent être aussi collectées à l'aide de questionnaires implémentés dans des applications pour *smartphone* ou par le recensement d'images de comportement [...] Si ces données peuvent aider à repérer plus facilement et rapidement les personnes nécessitant une prise en charge, le traitement ou la diffusion de ces informations peuvent également les exposer à des risques importants et posent d'importantes questions éthiques ».

1.3 Données relatives à des spécificités ethniques

Parmi les données de santé, il est possible de singulariser celles qui ont trait à l'ethnie réelle ou supposée de la personne. En effet, des maladies affectent particulièrement certaines populations, comme la drépanocytose qui touche principalement les populations originaires d'Afrique sub-saharienne et des Antilles. Il est donc fondamental de disposer d'informations concernant l'ethnie pour faire avancer la recherche, au risque, sinon, de négliger certaines maladies. Pourtant, la France entretient avec les données ethniques une relation très complexe qui lui est spécifique par contraste avec d'autres pays européens[18].

Aujourd'hui, le cadre règlementaire français qui s'appuie sur l'article 1er de la Constitution et sur l'article 6.1 de la Loi Informatique et Libertés[19] interdit la collecte de données ethniques, sauf pour une destination précise sous autorisation spéciale de la CNIL. On peut citer par exemple l'enquête TeO (Trajectoires et Origines) de l'Institut national d'études démographiques (INED) sur la question de la diversité des populations en France[20]. Dans le domaine de la recherche médicale, la CNIL autorise « les recherches nécessitant la réalisation d'un examen des caractéristiques génétiques » dans le cadre de ses

[17] Avis commun n° 141 du CCNE et n°4 du CNPEN, (2023), *Diagnostic Médical et Intelligence Artificielle : Enjeux Ethiques*, 58 p.

[18] Le Monde, « Statistiques ethniques : une situation contrastée en Europe », 05/02/2010.

[19] Voir : https://www.cnil.fr/fr/la-loi-informatique-et-libertes#article6 : « Il est interdit de traiter des données à caractère personnel qui révèlent la prétendue origine raciale ou l'origine ethnique [...] d'une personne physique ou de traiter des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique [...] »

[20] Voir le site de l'enquête : https://teo1.site.ined.fr/.

méthodologies de référence de 2018 sous certaines conditions[21]. Néanmoins cette place faite aux exceptions motivées par leur objet ne résout pas le problème posé par l'éventuelle disponibilité de ces données dans des PDS, car les usages qui en seraient faits ultérieurement ne sont pas connus *a priori*.

À ce propos, le compte-rendu de la « Mission d'information sur l'émergence et l'évolution des différentes formes de racisme et les réponses à y apporter »[22] souligne que :

> « la loi dite informatique et libertés et le RGPD apportent des garanties vis-à-vis des données dites « sensibles », dont celles qui révèlent l'origine raciale ou ethnique. Les considérants du RGPD indiquent que « l'utilisation de l'expression "origine raciale" dans le présent règlement n'implique pas que l'Union [Union Européenne] adhère à des théories tendant à établir l'existence de races humaines distinctes ». Dans la loi française, il est question de « prétendue origine raciale des personnes ».

Cette question est si complexe, même si on la restreint aux seules données de santé – au point qu'il est difficile de choisir le vocabulaire pour en parler, comme en témoigne le titre de cette sous partie -, qu'elle ne peut être traitée dans le cadre présent de notre réflexion sur les PDS. Nous pensons qu'il s'agit d'une question profonde et grave. Elle est inscrite au programme de travail du CCNE et du CNPEN qui produiront un avis sur la question dans un délai raisonnable.

2. Typologie des infrastructures de données de santé

Dans cet avis, nous nous intéressons aux structures rassemblant des données collectées ou permettant d'échanger ces données, et offrant des moyens de traitement de ces informations. Les organisations qui rassemblent des données de santé peuvent être privées ou publiques : il peut s'agir de laboratoires, d'hôpitaux, de cliniques, ou parfois d'acteurs en marge du parcours de soins. Les structures de données de santé sont appelées systèmes d'information de santé ou bases de données de santé (BDS), entrepôts de données de santé (EDS), ou plateformes de données de santé (PDS). Elles peuvent être locales avec un lieu de stockage physique unique (hub), organisées en réseau ou interrogées par l'intermédiaire d'une plateforme médiatrice. Les plus élaborées offrent des moyens de traitement (espace de calcul, logiciels de traitements numériques, algorithmes d'IA, etc.). Nous donnons tout d'abord une définition de ces différents types de structures de données de santé, avant de nous focaliser sur les plateformes de données de santé.

2.1 Bases de données de santé

Concernant les **bases de données de santé (BDS)**, nous repartons de la définition donnée dans l'avis 130 du CCNE[23] que nous précisons : une base de données est un ensemble structuré et organisé permettant la conservation de grandes quantités d'informations, portant sur un domaine spécialisé, afin d'en permettre l'exploitation. Cette définition rejoint celle de l'article premier de la directive 96/9/CE du Parlement européen et du Conseil du 11 mars 1996 concernant la protection juridique des bases de données, qui sont vues comme des « recueils d'œuvres de données ou d'autres éléments indépendants, disposés de manière systématique ou méthodique et individuellement accessibles par des

[21] CNIL, « Méthodologie de référence MR-001. Recherches dans le domaine de la santé avec recueil du consentement » [https://www.cnil.fr/fr/declaration/mr-001-recherches-dans-le-domaine-de-la-sante-avec-recueil-du-consentement].

[22] Mission d'information sur l'émergence et l'évolution des différentes formes de racisme et les réponses à y apporter, Compte rendu n° 47, 17 novembre 2020. [https://www.assemblee-nationale.fr/dyn/15/comptes-rendus/racisme/l15racisme2021047_compte-rendu#].

[23] CCNE, avis n°130 du 29 mai 2019, Données massives et santé : une nouvelle approche des enjeux éthiques, p. 64.

moyens électroniques ou d'une autre manière ». Notre définition insiste sur le caractère massif des bases de données.
Les données peuvent être intégrées dans une base de données telles quelles (éventuellement fournies par diverses sources extérieures) ou annotées (traitées par les gestionnaires de ces bases). Les utilisations principales d'une base de données sont l'accès au support pour retrouver, exploiter et diffuser des données ou l'interrogation pour faire par exemple des statistiques.
Les bases de données de santé sont issues pour certaines du programme de médicalisation des systèmes d'information (PMSI)[24], qui permet de « décrire de façon synthétique et standardisée l'activité médicale des établissements de santé. Il repose sur l'enregistrement de données médico-administratives normalisées dans un recueil standard d'information ». C'est le cas, par exemple, des bases de données des établissements hospitaliers qui recueillent et structurent les données sur les soins de suite et de réadaptation (PMSI-SSR[25]).

2.2 Entrepôts de données de santé

En informatique, un entrepôt de données est une infrastructure informatique qui rassemble en un lieu physique des données, pouvant être exprimées dans des formats variés, provenant de plusieurs sources souvent hétérogènes, et parfois de natures très différentes. Le caractère physique de l'entrepôt en fait un lieu matériel comprenant du hardware et des intervenants humains. Généralement, l'entrepôt de données informatique permet d'intégrer ces données hétérogènes selon un modèle unifié, de manière à faciliter pour l'utilisateur leur exploitation. À noter que l'expression **entrepôt de données de santé** (EDS) est employée même en l'absence d'un modèle unifié permettant une intégration informatique efficace, dès lors que le lien entre les différentes données des bases de données rassemblées dans l'entrepôt peut être établi. Par ailleurs, les EDS se couplent de plus en plus avec des plateformes de données (définies ci-dessous) pour pouvoir permettre *in situ* des traitements sur leurs données.

En 2019, la CNIL distingue formellement[26] une BDS, simple structure informatique qui rassemble des données de santé en vue d'une recherche, étude ou évaluation ponctuelle, et un EDS, une structure informatique qui permet de réaliser ultérieurement plusieurs traitements. La CNIL a adopté un référentiel[27] le 17 novembre 2021 sur les entrepôts, pour simplifier les procédures. Ce référentiel permet aux organismes voulant mettre en œuvre un EDS conforme au référentiel de ne pas solliciter d'autorisation préalable auprès de la CNIL. Il ne s'applique qu'aux EDS reposant sur l'exercice d'une mission d'intérêt public, au sens de l'article 6.1.e du RGPD. Dans ce référentiel, elle considère que « les EDS sont des bases de données destinées à être utilisées notamment à des fins de recherches, d'études ou d'évaluations dans le domaine de la santé ». On notera à cet égard que la réutilisation des données d'un EDS pour un projet particulier constitue un « traitement de données » à part entière au sens du RGPD ; chaque projet devra donc disposer d'un espace de traitement de données qui lui soit propre, cloisonné des autres projets sur l'EDS.

---

[24] Voir : https://solidarites-sante.gouv.fr/professionnels/gerer-un-etablissement-de-sante-medico-social/financement/financement-des-etablissements-de-sante-10795/financement-des-etablissements-de-sante-glossaire/article/programme-de-medicalisation-des-systemes-d-information-pmsi

[25] Voir : https://www.atih.sante.fr/presentation-pmsi-ssr

[26] CNIL, « Traitements de données de santé : comment faire la distinction entre un entrepôt et une recherche et quelles conséquences ? » [https://www.cnil.fr/fr/traitements-de-donnees-de-sante-comment-faire-la-distinction-entre-un-entrepot-et-une-recherche-et]

[27] CNIL, « La CNIL adopte un référentiel sur les entrepôts de données de santé » [https://www.cnil.fr/fr/la-cnil-adopte-un-referentiel-sur-les-entrepots-de-donnees-de-sante]

2.3 Plateformes de données de santé

Nous proposons de définir une **plateforme de données de santé** (PDS) comme un EDS qui offre de plus des services pour partager, traiter et analyser les données, tels que des logiciels et des capacités de calcul sur des serveurs de haute capacité. Les traitements algorithmiques peuvent être à base d'apprentissage machine et nécessitent très souvent des masses de données importantes. Sans prétendre à l'exhaustivité différents types de PDS françaises ou étrangères sont présentées en annexe (Annexe 4.1 à 4.8) depuis des BDS et EDS anciens qui se sont équipés de services jusqu'à des structures créées récemment pour mutualiser et accroître l'offre de service comme le GIP-PDS.

2.4 Cohortes

Les cohortes sont des cas particuliers de BDS qui constituent l'un des outils emblématiques utilisés par l'épidémiologie pour étudier la distribution des maladies et des invalidités dans les populations humaines ainsi que les influences qui déterminent cette distribution. Leur principe est de sélectionner un ensemble de personnes volontaires partageant éventuellement un certain nombre de caractéristiques communes, et de les suivre dans le temps à l'échelle individuelle afin d'identifier la survenue d'événements de santé d'intérêt. En France on comptait en 2019 plus de 250 études de cohorte. Mais certaines cohortes à l'échelle de dizaines ou centaines de milliers de personnes peuvent être mises à disposition de plusieurs projets de recherche et se constituer alors en véritables plateformes de données de santé. Nous en donnons deux exemples en annexe (Annexe 4.9 et 4.10) avec *Constances*, fondée sur une cohorte française, et *UK Biobank*, reposant sur une cohorte britannique.

2.5 Courtiers en informations (*data brokers*)

À côté des plateformes de données de santé, des entreprises se sont développées autour de la collecte d'informations personnelles, typiquement à travers des activités en ligne, pour organiser leur marché : ce sont les **courtiers en informations** ou ***data brokers.*** Ces données commercialisées peuvent être soit des données de santé d'usage primaire, mais en principe anonymisées (cf §I.3.2.2), ou bien des informations personnelles issues des traces laissées sur internet, telles que des messages sur les réseaux sociaux ou les forums, mais aussi à travers des applications de santé (IdO[28], traceurs, capteurs) ou de bien-être (fitness, sport). Cela peut aussi inclure des reçus de commandes de pharmacies en ligne, l'historique de téléconsultation ainsi que d'autres sources d'informations médicales publiques ou non. A cela peuvent s'ajouter des données de localisation et de fréquentation des lieux médicaux (cliniques, hôpitaux) ou de salles de sport. À noter qu'il peut être difficile de tracer l'origine de ces données. Un exemple de data broker en santé est l'entreprise IQVIA-France[29], spécialisée dans le marché des médicaments et qui a accumulé des données relatives aux tests antigéniques et vaccins anti-Covid depuis le début de la pandémie. Un autre exemple est l'entreprise Cegedim[30] qui développe et commercialise des bases de données et des logiciels dans le domaine de la santé.

2.6 Plateformes d'échange de données

Outre les courtiers en informations dont le principal objet de commercialiser l'accès à des données collectées, il faut mentionner d'autres acteurs commerciaux que sont les plateformes d'échange de données qui mettent en contact fournisseurs et acquéreurs de

[28] IdO : internet des objets (IoT en anglais, *Internet of Things*).

[29] Voir : https://iqvia.opendatasoft.com/pages/accueil/ et https://www.data.gouv.fr/fr/organizations/iqvia-france/.

[30] Voir : https://www.cegedim.fr/.

données en permettant leur échange dans un environnement sécurisé, sans jamais stocker ces données. Ces plateformes d'échange peuvent être administrées par des entreprises, comme Dawex, ou des coopératives, comme Salus Co-op en Espagne, une plateforme d'échange de données en libre accès, ou Healthbank, une initiative payante en Suisse, ou encore Doctolib en France. Ces exemples de plateforme d'échange de données de santé sont présentées en annexe (Annexes 4.11 à 4.14). Plusieurs d'entre elles seraient éligibles au rôle de « prestataire de service d'intermédiation de données » que prévoit le règlement européen sur la gouvernance des données (*Data Governance Act*) du 23 juin 2022 [31] (voir §III.4.2)

3. Pourquoi échanger, rassembler et traiter des données de santé massives ?

Les motifs qui ont conduit à la création du GIP-PDS qui s'appuie en particulier sur l'entrepôt de données du Système National des Données de Santé (SNDS) (voir annexes 4.1 et 4.2) illustrent la problématique de la création de PDS, dont certaines préexistaient.

3.1 Motifs de la création de plateformes de données de santé

3.1.1 Vers la médecine des 4P

La médecine est devant plusieurs révolutions qui devraient lui permettre de mieux répondre aux défis sanitaires dans nos sociétés. La première révolution concerne la médecine de précision qui consiste à développer des interventions ciblées prenant en compte le profil génétique et environnemental de la personne. La deuxième révolution est la possibilité de mieux cerner les facteurs de risques, de protection et de résilience qui vont permettre de prévenir et de prédire et diagnostiquer l'émergence de la maladie au sein de populations plus à risque de développer des maladies spécifiques. Enfin, la troisième révolution est le développement d'une médecine participative co-développée avec l'ensemble des personnes concernées (patients, aidants, personnels médical et administratif). Ces évolutions vers la médecine dite des 4 P - personnalisée, préventive, prédictive et participative - nécessitent de travailler sur des données à la fois à grande échelle et de très grande qualité.

3.1.2 Enjeux pour la recherche, l'intérêt public et les organisations privées

Le GIP-PDS a été créé en novembre 2019 pour faciliter le partage et l'utilisation des données de santé issues de sources très variées afin de favoriser la recherche dans ce domaine. Sa création fait suite aux préconisations du rapport Villani[32] sur l'intelligence artificielle (IA) qui prônait le partage des données de santé, et répond à la volonté du Président de la République de faire rayonner la France à l'international dans le domaine du numérique en santé.

De manière générale les PDS rassemblent en leur sein des données de santé d'origines très variées, généralement dispersées, et hétérogènes, pour faciliter leur exploitation conjointe. Les rassembler sur une plateforme permet de les rendre accessibles, de les standardiser, et rend possible leur traitement numérique avec des capacités de calcul ou des algorithmes d'IA élaborés comme ceux d'apprentissage profond (*deep learning*).

[31] Voir : https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R0868. Publié au Journal Officiel de l'Union Européenne le 23 juin 2022, le DGA entrera en vigueur en septembre 2023

[32] Villani C., (2018), *Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne*, 235 p. [https://www.vie-publique.fr/sites/default/files/rapport/pdf/184000159.pdf].

Le croisement et le traitement des données de santé très variées permettent de répondre à de multiples objectifs : faire avancer la recherche biomédicale et l'innovation, la prise de décision médicale (prescription, interprétation d'examens biologiques), les soins cliniques, la veille sanitaire et la matériovigilance (surveillance des dispositifs médicaux en conditions réelles), et enfin aider le pilotage du système de santé.

3.2 Protection de la personne, intérêt public et bien commun

3.2.1 Recherche scientifique et intérêt public

L'utilisation des données de santé pour l'intérêt public doit se réaliser en préservant la protection des données personnelles. Le RGPD préconise le consentement individuel sauf exceptions mentionnées ci-dessus (§ I.1.2). Toutefois, dans le cas de la recherche scientifique, les contours de certaines de ces définitions nécessitent d'être précisés. Comme le fait remarquer M. Shabani[33], le RGPD donne peu d'indications pour déterminer ce qui peut être considéré comme de la recherche scientifique d'intérêt public et général, notamment lorsque des entités commerciales sont impliquées. Alors que le RGPD considère la recherche soutenue par des fonds privés comme de la recherche scientifique, il ne distingue pas selon qu'elle vise le profit ou pas, pas plus qu'il n'indique si l'intérêt public prime sur les intérêts privés et commerciaux. Par ailleurs, l'interprétation de l'exception de la recherche (*research exemption*) dans les États relevant du RGPD fait apparaître une grande disparité[34]: 18 pays, dont la France, ont développé une réglementation spécifique sur la recherche et l'intérêt public, tandis que seuls 9 pays ont adopté des dispositions spécifiques dans le cas de recherches conduites par des organismes privés.

En France, l'article 66 de la Loi informatique et Liberté[35] dispose que « La garantie de normes élevées de qualité et de sécurité des soins de santé et des médicaments ou des dispositifs médicaux constitue une finalité d'intérêt public ». La CNIL précise ce qu'il faut entendre par mission d'intérêt public[36] : « La mission d'intérêt public est une des six bases légales prévues par le RGPD autorisant la mise en œuvre de traitements de données à caractère personnel. […] Cette base légale concerne donc en premier lieu les traitements mis en œuvre par les autorités publiques. Elle peut néanmoins autoriser la mise en œuvre de traitements par des organismes privés, dès lors qu'ils poursuivent une mission d'intérêt public ou sont dotés de prérogative de puissance publique », comme les établissements privés de santé chargés d'une mission de service public. Cette base légale dispense de la demande du consentement.

Dans le domaine de la santé, le GIP-PDS donne des indications aux porteurs de projets – qui peuvent être des industriels de la santé comme des assureurs – pour les aider à identifier si leur projet poursuit bien une finalité d'intérêt public, en se référant au SNDS[37] :

> « En plus de poursuivre une finalité d'intérêt public, les projets doivent vérifier les finalités propres au SNDS, à savoir contribuer (i) à l'information sur la santé, ainsi que sur l'offre de soins, la prise en charge médico-sociale et leur qualité ; (ii) à la définition,

[33] Shabani M., (2022), "Will the European Health Data Space change data sharing rules?", *Science*, vol. 375, Issue 6587.

[34] European commission, Assessment of the EU Member States' rules on health data in the light of GDPR, (2021):[https://health.ec.europa.eu/system/files/2021-02/ms_rules_health-data_en_0.pdf]

[35] Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés [https://www.legifrance.gouv.fr/loda/article_lc/LEGIARTI000038888793/].

[36] CNIL, « La mission d'intérêt public : dans quels cas fonder un traitement sur cette base légale ? » [https://www.cnil.fr/fr/les-bases-legales/mission-interet-public].

[37] Health Data Hub, « Qu'est-ce que l'intérêt public ? » [https://health-data-hub.fr/interet-public].

à la mise en œuvre et à l'évaluation des politiques de santé et de protection sociale ; (iii) à la connaissance des dépenses de santé, des dépenses d'assurance maladie et des dépenses médico- sociales ; (iv) à l'information des professionnels, des structures et des établissements de santé ou médico-sociaux sur leur activité ; (v) à la surveillance, à la veille et à la sécurité sanitaires ; (vi) à la recherche, aux études, à l'évaluation et à l'innovation dans les domaines de la santé et de la prise en charge médico-sociale ».

En outre, le GIP-PDS précise qu'il est interdit de procéder à un traitement « soit qui aurait pour objectif d'aboutir à prendre une décision à l'encontre d'une personne physique identifiée sur le fondement des données la concernant et figurant l'un de ces traitements, soit qui viserait la promotion en direction des professionnels de santé ou des établissements des produits de santé, ou l'exclusion de garanties des contrats d'assurance ou la modification de cotisations ou de primes d'assurance pour un individu ou un groupe d'individus ».

3.2.2 Marchandisation des données personnelles ou bien commun

Si l'entrée en vigueur du RGPD a clos le débat sur le statut juridique de la donnée à caractère personnel en Europe, en améliorant considérablement sa protection, il n'en demeure pas moins qu'il existe aussi, en Europe, un marché où s'échangent des données que l'on appelle parfois « médicales » pour les distinguer des données « de santé ». Ces données commercialisées doivent avoir été anonymisées (cf § I.4.6). Un courtier en informations ou *data broker* (voir § I.2.5) pourra ainsi vendre légalement par exemple mille radios de fracture du poignet, à la condition que celles-ci ne comportent aucune information permettant la réidentification.

D'autre part, le droit de propriété et de cession des données de santé (qui sont donc identifiantes) est pratiqué dans certains pays par le biais des mêmes courtiers en information. Pour les tenants de cette évolution, ce n'est pas tant le principe de la cession qui est au cœur du débat mais celui de la cession à titre onéreux. Reconnaître un droit de cession n'est pas en soi primordial puisque la donnée à caractère personnel peut être communiquée à des tiers. Mais accorder un droit de propriété sur les données à caractère personnel permettrait de reconnaître leur valeur financière. Leur cession pourrait alors faire l'objet d'une rémunération.

Beaucoup d'institutions se sont prononcées contre une telle évolution du droit, en soulignant l'ampleur de la tâche (définir un régime juridique nouveau) pour un faible intérêt sur le plan financier. La donnée d'un seul individu, à la différence de la donnée agrégée ou enrichie, n'a que peu de valeur marchande. De manière plus générale, on peut s'interroger d'un point de vue éthique sur la vente et donc la marchandisation d'une information qui est jusqu'ici considérée comme intimement liée à la personne et à son identité.

Reste la voie encore peu explorée et radicalement opposée qui est celle des biens communs, inspirée des travaux sur l'économie institutionnelle d'Elinor Ostrom (prix Nobel d'économie en 2009). Il s'agit de ressources communes dont la gestion repose sur la mise en place, par les utilisateurs, d'un système de règles socialement sanctionnées. Le logiciel libre est le meilleur exemple de bien commun dans le domaine numérique. Elinor Ostrom a développé une théorie des communs de la connaissance qui s'applique à la gestion de ressources d'ordre immatériel comme l'information. Traduire cette notion économique en notion juridique, alors qu'elle est inédite en France en tant que telle (à l'exception des biens de sections de commune), nécessiterait une création juridique qui pourrait s'inspirer de notions proches, notamment les choses communes régies par l'article 714 du Code civil. Une telle évolution paraît néanmoins plus adaptée au statut juridique des données

agrégées ou des réseaux de données que de la donnée elle-même. Cette évolution serait intellectuellement plaisante mais les différences concrètes qu'elle apporterait pour la gestion des données ne sont pas évidentes à envisager.

3.3 Exemples de projets portés par des plateformes de données de santé

Pour illustrer l'intérêt de rassembler sur des PDS des données de santé variées, nous présentons quelques exemples de projets qui utilisent soit le GIP-PDS présenté en annexe (Annexe 4.2), soit des grandes cohortes comme Constances (Annexe 4.9) ou *UK Biobank* (Annexe 4.10) sans autre spécificité que d'avoir été facilement identifiables et présentables par les membres de ce groupe de travail. A noter que pour ce qui concerne le GIP-PDS on ne dispose pas encore de projets achevés dont on pourrait évaluer les résultats. Pour chaque projet les données sont traitées dans un espace spécifique au projet qui ne contient que les données nécessaires. Le responsable du projet a accès à distance aux données et ne peut pas les récupérer.

3.3.1 HYDRO (GIP-PDS)

Le projet HYDRO[38] (Développement et validation d'algorithmes de prédiction des crises d'Insuffisance cardiaque chez les patients porteurs d'implants connectés) est porté par la société privée Implicity[39]. Il a reçu toutes les autorisations et a commencé en octobre 2021 pour une durée de cinq ans. Il apparie des données du SNDS et des données de la plateforme d'Implicity utilisée par les cardiologues pour le télésuivi du patient, selon un processus[40] validé par le GIP-PDS.

3.3.2 Glucocorticoïdes (GIP-PDS)

Le projet Glucocorticoïdes[41] est intéressant, même s'il ne fait pas d'appariement sur les données, car c'est un projet international initié par l'Agence Européenne du Médicament (EMA) qui implique sept États. Le responsable du traitement est la multinationale américaine IQVIA, spécialiste des données de santé. L'étude a pour but d'améliorer la prise en charge des patients atteints du SARS-CoV-2 ou présentant une suspicion d'atteinte par cette maladie. Il s'agit d'utiliser des données de santé publique en recourant aux données du SNDS afin de déterminer les patients pouvant le plus bénéficier de l'utilisation de glucocorticoïdes, et choisir les meilleures molécules, ainsi que le meilleur dosage et le meilleur moment d'administration. L'étude a pour objectif de décrire les schémas d'utilisation des glucocorticoïdes par voie systémique, ainsi que les événements indésirables associés à ces médicaments à partir d'une cohorte de patients SARS-CoV-2 au cours des années 2019 et 2020. Elle prévoit également de valider la faisabilité d'études européennes sur les traitements du SARS-CoV-2 en utilisant des données mises au format international OMOP[42] par les ingénieurs de la PDS, à la demande de l'EMA.

---

[38] Health Data Hub, « HYDRO : Développement et validation d'algorithmes de prédiction des crises d'Insuffisance cardiaque chez les patients porteurs d'implants connectés » [https://www.health-data-hub.fr/projets/hydro-developpement-et-validation-dalgorithmes-de-prediction-des-crises-dinsuffisance].

[39] Voir le site : https://www.implicity.com/.

[40] Voir : https://www.health-data-hub.fr/sites/default/files/2021-10/Infographies%20Projet%20HYDRO.pdf

[41] Health Data Hub, « Utilisation des glucocorticoïdes par voie systémique dans le traitement de la COVID-19 et risques d'événements indésirables » [https://health-data-hub.fr/projets/utilisation-des-glucocorticoides-par-voie-systemique-dans-le-traitement-de-la-covid-19-et].

[42] Voir : https://www.ohdsi.org/data-standardization/the-common-data-model/.

3.3.3 REXETRIS (GIP-PDS)

Le projet REXETRIS[43] (Relations EXposition - Effet à long terme chez le Transplanté Rénal des médicaments ImmunoSuppresseurs) est porté par le CHU de Limoges et soutenu par la société Optim'Care. Il vise à améliorer le suivi des patients transplantés rénaux en proposant de nouvelles modalités d'optimisation des traitements immunosuppresseurs plus ciblés. Pour cela, il étudie une cohorte rétrospective sur des données issues de trois bases de données : CRISTAL de l'Agence de la Biomédecine, SNDS (base CEPIDC) de la CNAMTS et ABIS du CHU de Limoges, base utilisée dans le cadre de la prise en charge des patients. Le projet envisage deux appariements afin de constituer une base pseudonymisée réunissant l'ensemble des patients transplantés rénaux suivis dans un centre de transplantation français depuis 2005 : entre ABIS et CRISTAL, et entre CRISTAL et SNDS. Il a reçu toutes les autorisations de traitement et a commencé en juillet 2021 pour une durée de cinq ans. La conservation des données une fois le projet finalisé est de sept ans.

3.3.4 Pathologies chroniques (Constances)

La PDS Constances fondée sur une grande cohorte française a permis de mener à bien plusieurs projets de recherche sur des pathologies chroniques dont par exemple : l'activité sexuelle des femmes diabétiques, la bronchopneumopathie chronique obstructive et l'infection au VIH, ou encore les morbidités liées à la consommation d'alcool.[44]

3.3.5 Neuroanatomie du cerveau (UK Biobank)

La PDS UK Biobank, fondée sur une grande cohorte britannique a permis d'obtenir des résultats significatifs sur la neuroanatomie du cerveau en reliant des données épidémiologiques et génétiques à des données d'imagerie par résonance magnétique (IRM) du cerveau, du cœur et de l'abdomen auprès de 100 000 participants. L'analyse combinée des données de génotypage et d'IRM cérébrales a permis, entre autres, d'estimer l'héritabilité des différences interindividuelles[45, 46].

3.4 Trois premières questions éthiques liées à la collecte de données massives

Le volume des données rassemblées dans les PDS doit être suffisamment grand et de qualité, pour pouvoir permettre des traitements intéressants, tels que ceux effectués avec de l'apprentissage machine, qu'il s'agisse des données pour un usage primaire, recueillies auprès des patients, comme celles du SNDS ou de *Mon Espace Santé*, ou des données de santé pour un usage secondaire issues de projets de recherche, dont la qualité dépend des données pour un usage primaire, des algorithmes de traitement et des opérations de « curation » qu'elles ont subies (c'est-à-dire le travail de maintenance et de nettoyage). Toutefois, il convient de s'interroger sur le bien-fondé de vouloir stocker massivement pour constituer des bases de données les plus couvrantes possibles, selon ce qu'on peut appeler un principe de précaution inversé. L'avènement des *Big Data* a en effet donné lieu

[43] Health Data Hub, « REXETRIS : Relations EXposition - Effet à long terme chez le Transplanté Rénal des médicaments ImmunoSuppresseurs » [https://www.health-data-hub.fr/projets/rexetris-relations-exposition-effet-long-terme-chez-le-transplante-renal-des-medicaments].

[44] Voir : https://www.constances.fr/projets-termines.

[45] Biton, A. *et al.*, (2020), "Polygenic Architecture of Human Neuroanatomical Diversity", *Cereb. Cortex N. Y. N 1991,* **30**, 2307–2320.

[46] Elliott, L. T. *et al.*, (2018), "Genome-wide association studies of brain imaging phenotypes in UK Biobank", *Nature* **562**, 210–216.

à une véritable boulimie de données qui soulève trois tensions éthiques, qui nous interpellent sur le modèle épistémique des *Big Data* en santé[47].

L'efficacité et la pertinence des traitements effectués sur les PDS dépendent du volume mais aussi de la qualité des données utilisées. Comme le souligne le rapport ITF – Sopra-Steria Next, l'une des dimensions principales de la qualité des données est leur représentativité[48]. Ainsi, selon l'association Filière Intelligence Artificielle et Cancers,[49] plus les données proviennent de sources variées, plus les risques de biais sont minorés dans la recherche sur le cancer. La qualité se retrouve intrinsèquement liée à la diversité des données.

Mais les utilisateurs de PDS ne peuvent trouver que ce qui est stocké dans les PDS (à moins de solliciter de nouvelles données pour leur étude, après avoir recueilli le consentement des personnes concernées). Il y a alors une tension entre la représentativité des données (pour éviter tout biais), et le respect de l'autonomie des personnes qui peuvent ou non donner leur consentement (voir § III.1 à § III.4) pour que leurs données soient utilisées et stockées. Ainsi la diversité et la représentativité des données ne peuvent pas être toujours assurées. Donc il faut que les concepteurs de PDS spécifient de façon précise ce qui est stocké dans la PDS pour une utilisation éclairée des données de la PDS (voir Recommandation n°1).

Il revient aussi aux utilisateurs des PDS de bien choisir leur ensemble de données, eu égard aux finalités de leur projet et de remédier aux biais éventuels, comme, le fait, par exemple, la plateforme d'échange de données Dawex[50] qui propose des outils d'échantillonnage pour générer automatiquement des exemples de données représentatifs sur la base d'algorithmes, afin d'éviter tout biais[51]. S'il s'avère impossible d'éviter les biais, une méthode de pondération pour les prendre en compte peut être mise en œuvre comme le fait par exemple *UK BioBank* [52] (voir Recommandation n° 1).

La nécessité de stocker des données massives pour utiliser des outils numériques efficaces sur les PDS tels que des systèmes d'apprentissage automatique entre en tension avec le principe de minimisation de la collecte, du stockage et de la durée de conservation, tel que prescrit par le RGPD. De plus, le volume de données susceptibles d'être piratées en raison de failles de sécurité ou d'utilisations malveillantes augmente avec le volume et la durée de stockage des données. Ceci présente donc un risque pour la protection des données personnelles sensibles et le respect de la vie privée.

On ne rappellera donc jamais trop la nécessité de respecter le principe de proportionnalité de la collecte au regard des finalités précises (voir principe de minimisation du RGPD), y compris dans le cadre de la collecte des données pour la recherche. Néanmoins, la minimisation de la durée de conservation des données de santé collectées en situation de

---

47 Dans cet avis, nous n'abordons pas les questions éthiques soulevées par la numérisation croissante des données et outils de santé. Nous nous limitons ici aux enjeux d'éthique soulevés par le stockage et l'échange de données massives de santé à travers les plateformes de données de santé.

48 Rapport HTF – Sopra-Steria Next, (2022), « Data-altruisme, une initiative européenne. Les données au service de l'intérêt général », Rapport Human Technology Foundation, et Exploratoire Sopra-Steria Next., p 26.

49 Voir : https://filiere-ia.fr/.

50 Voir : https://www.dawex.com/ et annexe 4.11

51 Voir : https://digital-strategy.ec.europa.eu/en/policies/data-governance-act-explained#ecl-inpage-l4ihmeih

52 Voir : https://www.medrxiv.org/content/10.1101/2022.05.16.22275048v1.full

crise sanitaire qui relève de l'appréciation du législateur doit prendre en compte les besoins de la recherche (voir Recommandation n°2).

Enfin, même si l'impact du numérique sur l'environnement, en particulier l'impact des services *cloud* (informatique en nuage) semble difficilement évaluable, la maximisation du stockage des données de santé pour améliorer les résultats à portée thérapeutique en vue du bien commun, va à l'encontre de la sobriété numérique, qui vise à limiter le stockage et les calculs numériques pour limiter l'impact sur l'environnement. Il convient donc d'évaluer l'impact environnemental des PDS actuelles et en constitution et de chercher à le minimiser, en particulier dans leurs usages pour la recherche, comme le recommande un avis récent du Comité d'éthique du CNRS[53] (voir Recommandation n°3).

4. Principes fondamentaux de constitution des architectures

Nous nous intéressons dans cette section aux enjeux d'éthique liés aux solutions informatiques choisies pour la conception des PDS. Les enjeux d'éthique liés aux modes de gouvernance et aux modèles économiques sont abordés dans les sections suivantes.

Dès la conception de l'architecture d'une PDS, il convient de respecter trois principes techniques fondamentaux qui ont des impacts d'ordre éthique : sécurité, interopérabilité et portabilité. Ces trois principes sont au cœur du RGPD.

4.1 Sécurité

La **sécurité** des infrastructures et des protocoles est une propriété technique cruciale exigée pour les plateformes de données en général, et en particulier pour les PDS, en raison du caractère sensible des données de santé. Elle concerne :

- les conditions de stockage dans les PDS, pour assurer l'intégrité des données (aucune personne non autorisée ne peut les modifier) et leur confidentialité (personne ne peut les extraire sans y être autorisé) ;
- les conditions de transfert des données, que ce soit lors de l'alimentation de la PDS (*upload*) à partir de sources disparates ou lors de l'utilisation des données pour effectuer des calculs hors site (*download*) ou échanger avec d'autres PDS.

La possibilité de violation des contenus des PDS, suite à des failles de sécurité ou d'utilisations malveillantes, soulève trois enjeux d'éthique : confidentialité, non malveillance et souveraineté.

Rappelons l'existence d'une solution publique, le CASD (Centre d'accès sécurisé aux données) [54], qui héberge des données très sensibles au sens du RGPD d'opérateurs publics et qui est un hub d'accès sécurisé aux données. Il peut donc être utilisé pour les données biomédicales. La cohorte *Constances* (voir annexe Annexe 4.9) développée par l'Inserm a recours au CASD pour ses projets de recherche par le biais de bulles sécurisées.

4.2 Interopérabilité

L'**interopérabilité** vise la possibilité d'échanger des données de santé entre diverses PDS. Elle favorise les échanges mais doit faire face à des problèmes de sécurité intrinsèques à l'échange de données entre PDS qui ne reposent pas sur les mêmes infrastructures et

[53] COMETS – Comité d'éthique du CNRS, « Intégrer les enjeux environnementaux à la conduite de la recherche – Une responsabilité éthique », Avis n° 2022-43, 5 décembre 2022. Voir : https://comite-ethique.cnrs.fr/avis-du-comets-integrer-les-enjeux-environnementaux-a-la-conduite-de-la-recherche-une-responsabilite-ethique/

[54] Voir : https://www.casd.eu/ et annexe 4.5.

n'utilisent pas les mêmes logiciels. Selon une analyse de l'association Gaia-X, les manques d'interopérabilité, de portabilité (voir § I.3.4) et de souveraineté des données (voir § II.1), sont les principales raisons qui semblent avoir empêché une adoption plus rapide du *cloud computing* en Europe[55].

L'interopérabilité suppose un effort de **standardisation** des systèmes de santé par la mise en place de référentiels, tant au niveau des systèmes d'information, des logiciels de traitement des données, qu'au niveau des formats pour la représentation des données.[56]

En ce qui concerne le *Ouest Data Hub* (ODH), l'interopérabilité par conception a été visée pour pouvoir communiquer de manière efficace avec les six EDS du réseau HUGO en s'appuyant sur une technologie commune.

De même, l'Agence du Numérique en Santé (ANS), en vue de la création de *Mon espace santé* pour tous et son utilisation par tous les médecins, a défini des référentiels d'interopérabilité qui facilitent la numérisation et l'homogénéisation des informations échangées, et a commencé par modifier les logiciels de formatage et stockage des données, pour anticiper l'échange des données.

L'interopérabilité passe aussi par l'élaboration de terminologies de santé. Concernant les données en oncologie, citons l'initiative du consortium interSIRIC OSIRIS, pour fédérer les bases de données spécialisées dans ce champ, qui a donné lieu à un modèle standardisé dynamique de représentation des données en oncologie[57].

Signalons aussi le projet de recherche-innovation Oncolab[58], lancé le 1er juillet 2022 par un consortium public-privé français, associant les sociétés Arkhn et Owkin, l'institut de recherche INRIA et des hôpitaux spécialisés dans le traitement du cancer : IUCT-Oncopole, Instituts Curie à Paris et Bergonié à Bordeaux, ainsi que le CHU de Toulouse. Ce projet vise à normaliser et à standardiser l'accès aux données de soins de santé pour la recherche en oncologie.

Un modèle commun de données OMOP-CDM (*Observational medical outcomes partnership – Common Data Model*)[59] a pour objectif **l'interopérabilité** entre les différentes bases d'analyse en santé, qu'elles soient cliniques ou médico-administratives. Il résulte d'un partenariat public-privé appelé OMOP (*Observational Medical Outcomes Partnership*), présidé par la *Food and Drug Administration* (FDA) aux États-Unis et financé par un consortium de sociétés pharmaceutiques créé en 2008 pour cinq ans. Il est utilisé par le GIP-PDS.

Le standard de communication FHIR (*Fast Healthcare Interoperability Resources*)[60] est un standard décrivant des formats de données et d'autres éléments ainsi qu'une interface de programmation applicative pour les échanges des informations médicales. Ce standard a été développé par *Health Level Seven International* (HL7), organisation à but non lucratif dédiée au développement de l'interopérabilité des données de santé et la standardisation du protocole d'échanges médicaux.

---

55 Tardieu H. et al., (2022), "Compliance, and consequences on the labeling framework of Gaia-X – Voir : https://gaia-x.eu/wp-content/uploads/2022/07/Gaia-X-Compliance-Document_Final_f.pdf

56 Voir par exemple le dossier de l'Inserm : « Big data en santé : des défis techniques et éthiques à relever » publié le 27/06/2022 [https://www.inserm.fr/dossier/big-data-en-sante/].

57 Voir : https://siric-brio.com/premiers-resultats-du-consortium-intersiric-osiris/.

58 Voir : https://www.bioworld.com/articles/520331-consortium-launches-oncolab-to-standardize-access-to-oncology-data?v=preview.

59 Voir : https://www.ohdsi.org/data-standardization/the-common-data-model/.

60 Voir : https://fr.wikipedia.org/wiki/Fast_Healthcare_Interoperability_Resources

Concernant les données génomiques, l'Alliance globale pour la génomique et la santé (*Global Alliance for Genomics and Health*, GA4GH[61]) propose divers standards ouverts pour le partage des données biomédicales. [62] [63]

Toutes ces observations ne font que renforcer l'enjeu de l'interopérabilité des PDS et il est souhaitable que la France s'implique davantage dans l'effort européen d'élaboration de standards et de normes pour formater et structurer les données de santé[64] (voir Recommandation n°4).

4.3 Réversibilité et portabilité

La **réversibilité** est la possibilité de changer de prestataire (fournisseur d'un service d'hébergement) ou de ré-internaliser les bases de données et les traitements associés, avec un coût minimal en termes d'applications, de données et d'infrastructures. Le GIP-PDS travaille à la réversibilité depuis sa préfiguration en 2019. S'il a choisi la solution Cloud Azure de Microsoft, seule disponible à court terme avec la certification « Hébergeur de Données de Santé » (HDS) et permettant d'intégrer les fonctionnalités et certifications nécessaires au niveau de sécurité requis, l'objectif de réversibilité est inscrit dans sa feuille de route depuis 2022[65].Or la réversibilité n'est permise que par la **portabilité** des données et des programmes informatiques.

La portabilité des données est définie dans le RGPD[66] : « Les personnes concernées ont le droit de recevoir les données à caractère personnel les concernant qu'elles ont fournies à un responsable du traitement, dans un format structuré, couramment utilisé et lisible par machine, et ont le droit de transmettre ces données à un autre responsable du traitement sans que le responsable du traitement auquel les données à caractère personnel ont été communiquées y fasse obstacle, lorsque le traitement est fondé sur le consentement (…), ou sur un contrat (…) et le traitement est effectué à l'aide de procédés automatisés. »[67] La CNIL précise que les données portables, recueillies avec l'accord de la personne concernée ou dans le cadre d'un contrat, doivent être fournies « dans un format structuré, couramment utilisé et lisible par une machine. Cela veut dire que l'organisme doit proposer des formats de données adaptés au type de données concernées, en privilégiant des formats ouverts, interopérables » [68].

Quant à la portabilité d'un programme informatique c'est sa capacité à pouvoir être adapté plus ou moins facilement en vue de fonctionner dans différents environnements d'exécution. Les différences peuvent porter sur l'environnement matériel (processeur) comme sur l'environnement logiciel (système d'exploitation). La différence d'environnement peut également porter sur une combinaison des deux éléments. C'est le cas, par exemple dans les domaines de l'informatique embarquée, des super calculateurs et des machines virtuelles.

---

61 Voir : https://www.ga4gh.org/.

62 Voir : https://www.ga4gh.org/genomic-data-toolkit/.

63 Rehm H.L. et al., (2021), "GA4GH: International policies and standards for data sharing across genomic research and healthcare", *Cell Genom*, 10; 1(2): 100029.

64 Voir : https://www.snomed.org/news-and-events/articles/EU-drives-standardized-terminology-funding-program.

65 Voir : https://www.health-data-hub.fr/sites/default/files/2022-01/HDH_Feuille_De_Route_2022_0.pdf.

66 Article 20 du chapitre 3 du RGPD : https://rgpd.com/fr/apercu/chapitre-3-droits-de-la-personne-concernee/article-20-droit-a-la-portabilite-des-donnees/

67 Les conditions sont : « lorsque le traitement est fondé sur le consentement en application de l'article 6, § 1, point a), ou de l'article 9, § 2, point a), ou sur un contrat en application de l'article 6, § 1, point b).

68 Voir : https://www.cnil.fr/fr/le-droit-la-portabilite-obtenir-et-reutiliser-une-copie-de-vos-donnees

La portabilité des algorithmes et des programmes utilisés sur les plateformes traitant des données de santé est un enjeu d'éthique important en ce qui concerne la reproductibilité des calculs ainsi que la capacité à bénéficier des compétences et efforts développés sur une plateforme pour bénéficier au plus grand nombre. Elle touche aussi aux questions de transparence et d'explicabilité, et elle peut grandement bénéficier d'approches issues du logiciel libre (*open source*) (voir § I.4.5). Cette portabilité est un sujet complexe et ancien[69] mais qui prend une importance nouvelle dans les conditions actuelles de souveraineté numérique. Elle dépend en effet des architectures matérielles et logicielles propres à chaque plateforme, dans un contexte d'utilisation d'algorithmes complexes de visualisation, d'apprentissage machine, de gestion et d'analyse de bases de données, d'utilisation de réseaux informatique et de cybersécurité. Cette portabilité est encore un sujet de recherche scientifique et technologique ainsi qu'un enjeu de standardisation. Elle doit être prise en compte depuis la conception des algorithmes jusqu'à leur programmation sur différentes architectures matérielles. Ce dernier point est délicat car les machines utilisées mettent en œuvre des architectures qui peuvent être très différentes les unes des autres. Il faut en particulier distinguer les situations dans lesquelles les programmes sont utilisés sur des machines standards (typiquement des stations de travail individuelles) de celles où des supercalculateurs sont mis en œuvre.

L'enjeu de la portabilité des données, des algorithmes et des programmes est donc stratégique pour la réversibilité des contrats sur les PDS, comme pour l'interopérabilité (voir Recommandation n° 4).

4.4 Architectures centralisées ou non

Si toutes les PDS existantes visent généralement à respecter les quatre principes discutés ci-dessus, elles se distinguent – entre autres – par la manière dont elles se sont constituées à partir de bases de données de santé existantes, ce qui a en particulier des conséquences d'ordre éthique.

Le GIP-PDS [70] a fait le choix d'une plateforme centralisée. Il s'est constitué en 2019 en s'appuyant sur le Système National de Données de Santé (SNDS) et en copiant dans sa plateforme centralisée et sécurisée les bases de données inscrites dans son catalogue. Le GIP-PDS peut aussi accueillir des données venant de sources non référencées au catalogue. Le stockage des données se fait dans le *cloud* Microsoft Azure [71], ce qui a suscité de vives controverses sur le respect de la souveraineté des données (voir § II.1.3).

Le *Ouest Data Hub* (ODH)[72] est un hub interrégional créé lui aussi en 2019 mais qui a adopté dès le départ une vision plus décentralisée et collaborative, en choisissant de mettre en réseau les EDS de plusieurs centres hospitaliers, et en s'appuyant sur leurs expertises locales en données cliniques et sur leurs cliniciens. Ces EDS ont été développés de la manière la plus homogène possible, en utilisant une technologie commune développée au sein de l'Inserm, en lien avec le CHU de Rennes afin de réaliser l'intégration des données en s'appuyant sur un partenariat public-privé avec la société Enovacom. Seules les données nécessaires aux projets sont stockées sur l'infrastructure de la PDS

[69] Voir : https://www.cigref.fr/archives/histoire-cigref/wp-content/uploads/2017/02/CIGREF-1977-portabilite-applications-informatiques.pdf.

[70] Voir annexe 4.2 et https://www.health-data-hub.fr/.

[71] L'Agence Nationale de la Sécurité des Systèmes d'Information (ANSSI) a développé un référentiel d'exigences pour les prestataires de services d'informatique en nuage SecNumCloud[71] et fourni une liste de prestataires qualifiés selon ce référentiel[71]. Parmi les prestataires agréés, on trouve deux français : OVH et Outscale SAS.

[72] Voir annexe 4.3 et https://www.chu-hugo.fr/accueil/projets/Ouest-DataHub/

qui est hébergée au sein du CHU de Nantes, et reconnue Hébergeur de Données de Santé (HDS).

Il est impossible actuellement d'évaluer intégralement les avantages et inconvénients des approches centralisées (avec stockage dans un *cloud*) par rapport aux approches non centralisées (hub interrégional avec infrastructure de la PDS hébergée en local dans un CHU) en termes de **sécurité** et de **confidentialité** des données.

Le fait de stocker les données en un lieu unique permet de n'avoir à surveiller qu'un seul coffre-fort, mais si ce coffre-fort est percé, la perte des données est plus conséquente que dans une approche décentralisée où l'attaque d'un coffre-fort parmi plusieurs concernerait moins de données. Il faudrait réfléchir à des études de simulation de failles de sécurité pour pouvoir mieux apprécier les risques de chacune de ces approches (voir Recommandation n°5).

Une approche décentralisée comme celle du ODH permet de s'appuyer sur des systèmes d'information existants (les six EDS des CHU de la région Ouest). Cela nécessite de prévoir une technologie commune pour le développement des EDS locaux, comme l'a fait le ODH, mais cela permet de prendre en compte et **respecter** le travail sur les systèmes d'information des hôpitaux locaux et de leur laisser une certaine **autonomie** dans la gestion de leurs entrepôts. De plus, la constitution de l'ODH s'est accompagnée du lancement de projets multicentriques, mettant en évidence la nécessité de partager les données entre CHU, ce qui a suscité un cercle vertueux et incité les cliniciens à adhérer au projet de ODH (voir Recommandation n° 6).

Concernant l'utilisation des PDS, une approche centralisée qui met à disposition dans un même lieu toutes les données nécessaires pour des traitements algorithmiques semble plus favorable à l'exploitation des données par des algorithmes d'apprentissage machine. Les tenants d'une approche décentralisée ou fédérée misent sur les avancées de la recherche et l'innovation dans le domaine de l'Intelligence artificielle fédérée (*federated Artificial Intelligence*) (voir Recommandation n° 5).

Dans les deux types d'approches, le coût des infrastructures, de la recherche et de l'innovation associées, ainsi que le besoin de ressources humaines, sont considérables et doivent être accompagnés.

4.5 Promotion de solutions techniques ouvertes

Comme le souligne la Haute Autorité de Santé (HAS), « la question de la propriété privée des schémas de données, notamment dans le cas de gestionnaires de dossiers patients informatisés, est une difficulté majeure »[73], sans compter que des liens d'intérêts et des problèmes déontologiques sont susceptibles d'apparaître[74]. En effet, ces opérateurs privés favorisent souvent l'utilisation de licences d'utilisation propriétaires et un code source fermé, ce qui pose aussi plusieurs problèmes, en particulier un manque de transparence et une dépendance forte vis-à-vis de l'éditeur dans le cadre de la maintenance et d'adaptation à de nouveaux produits.

---

[73] Voir : https://www.has-sante.fr/jcms/p_3386076/fr/entrepots-de-donnees-de-sante-hospitaliers-la-has-publie-un-panorama-inedit-en-france, p. 17. https://www.has-sante.fr/upload/docs/application/pdf/2022-11/rapport_entrepots_donnes_sante_hospitaliers.pdf page 28 ??

[74] Voir : https://www.lemonde.fr/planete/article/2019/12/24/donnees-de-sante-conflit-d-interets-au-c-ur-de-la-nouvelle-plate-forme_6023918_3244.html

C'est la raison pour laquelle il nous semble essentiel de promouvoir le développement de solutions ouvertes et de technologies *open source* pour les PDS comme dans d'autres domaines. Elles offrent une transparence du code par défaut et permettent à tous les acteurs de s'en emparer légalement. Leur réputation est parfois encore ternie parce que supposées moins fiables, mais de nombreuses expériences ont montré le contraire. En outre l'*open source* est un facteur important d'attractivité pour le recrutement de talents souvent attachés à ce modèle[75]. C'est la raison pour laquelle nous faisons nôtre la recommandation de la HAS prônant de favoriser l'émergence de plateformes technologiques ouvertes[76] (voir Recommandation n°7).

4.6 Pseudonymisation et anonymisation des données

La plupart des PDS recourent à la pseudonymisation des données de santé qu'elles recueillent. Ce faisant, ces données sensibles restent des données personnelles qui relèvent du RGPD. L'anonymisation des données serait une meilleure garantie de respect de la vie privée des patients. Cependant, les possibilités de croiser les données de santé avec d'autres bases sont de plus en plus nombreuses et la réidentification est de plus en plus souvent réalisable, notamment dans le cas de maladies rares, ce qui questionne la possibilité même de l'anonymisation. Par ailleurs, dans un certain nombre de cas, il peut être intéressant de retrouver les patients dont on analyse les données. Il y a donc une tension entre anonymisation des données pour le respect de la vie privée des patients et préservation de leur identité pour un meilleur soin.

Concernant les PDS recueillant des données génomiques, la situation est encore plus délicate car les données sont fortement identifiantes et ne concernent pas seulement la personne mais toute sa famille, tout en étant fondamentales pour développer la médecine de précision. À ce sujet, Bonomi et co-auteurs[77] font le point sur les principales menaces pour la vie privée dans le recueil et l'utilisation des donnés génomiques et les techniques de protection existantes ou à développer.

Le rapport ITF-Sopra-Steria Next[78] rappelle l'initiative non couronnée de succès de l'institut Robert Koch qui a mis en place en Allemagne, en avril 2020, dans les premiers mois de la pandémie, l'application Corona Data Donation. Avec cette application, la population était invitée à partager ses données de santé, notamment les symptômes liés à la contamination par le virus, et la pseudonymisation des données était garantie. Un an plus tard, le Commissaire fédéral allemand à la protection et la liberté d'information constatait que seulement un million d'utilisateurs avaient adopté cette application, expliquant cette faible participation par l'ambiguïté de la notion de « don des données et le choix de la pseudonymisation au lieu de l'anonymat total ». Ceci incite à développer des méthodes d'anonymisation pour mettre en confiance les patients susceptibles de confier leurs données de santé sur une PDS. Parallèlement à la recherche de nouvelles techniques d'anonymisation, des recherches se tournent vers d'autres voies, telles que l'exploration

---

[75] Alcaras, Gabriel. « Des logiciels libres au contrôle du code. L'industrialisation de l'écriture informatique ». Thèse de doctorat, Paris, EHESS, 2022. https://www.theses.fr/s341603.

[76] Voir : https://www.has-sante.fr/upload/docs/application/pdf/2022-11/rapport_entrepots_donnes_sante_hospitaliers.pdf, page 30.

[77] Bonomi L., Huang Y., & Ohno-Machado L., (2020), "Privacy Challenges and Research Opportunities for Genomic Data Sharing", Nat. Genet 52, 646–654.

[78] Rapport HTF – Sopra-Steria Next, (2022), « Data-altruisme, une initiative européenne. Les données au service de l'intérêt général », Rapport Human Technology Foundation, et Exploratoire Sopra-Steria Next., p 28.

des techniques homomorphes[79] ou la conception d'algorithmes de génération de données de synthèse anonymes – avatars – à partir de données personnelles[80] (voir Recommandation n° 8). Mais il n'existe pas encore actuellement de procédure d'anonymisation certifiée, en particulier par la CNIL.

4.7 Bonnes pratiques

Des travaux récents[81] font le point sur les approches adoptées par les scientifiques pour partager les données qu'ils utilisent. Il ressort de cette étude que, généralement, les scientifiques ont des démarches conformes aux principes éthiques et juridiques, mais qu'ils déplorent un manque de procédures pratiques qui garantiraient par défaut une collecte et un partage des données conformes à l'éthique et à la loi. Un ensemble de solutions technologiques permettant une éthique de conception (*ethics by design*) des plateformes de données de santé est souhaitée.

5. Recommandations

Sur la qualité et partage des données

- **Recommandation n°1** : Expliciter la nature et l'origine des données personnelles de santé rassemblées dans les PDS, en distinguant leurs usages primaires et leurs usages secondaires et, pour un projet de recherche donné, utiliser des ensembles de données non biaisées, ou, lorsque cela n'est pas possible, tenir compte de ces biais dans leur analyse, par exemple, par des méthodes de pondération.
- **Recommandation n°2 :** Veiller à ce que la durée de conservation des données de santé publique collectées soit bien calibrée par rapport aux exigences de la recherche sans négliger la nécessaire protection des données personnelles.

Sur l'impact environnemental des PDS

- **Recommandation n°3** : Evaluer l'impact environnemental des PDS et viser leur sobriété énergétique par des choix appropriés de stockage des données, d'architecture, et de modes de fonctionnement.

Sur l'architecture des PDS :

- **Recommandation n°4** : Demander aux pouvoirs publics de s'impliquer davantage dans l'élaboration de standards et de normes pour formater et structurer les données de santé afin de favoriser de meilleures portabilité et interopérabilité des PDS.
- **Recommandation n°5** : Mener des études d'évaluation comparatives entre les approches centralisées et les approches décentralisées des PDS et sur leurs combinaisons, pour assurer une gestion sécurisée des données de santé. Encourager les innovations en Intelligence Artificielle fédérée pour informer le débat entre architectures centralisées et décentralisées.
- **Recommandation n°6 :** Choisir des solutions d'architecture de PDS qui respectent les écosystèmes locaux et tiennent compte de projets de recherche multicentriques

[79] Gentry C., (2009), "A fully homomorphic encryption scheme", Thèse de doctorat, Stanford University.
[80] Voir, par exemple : https://www.larevuedudigital.com/anonymisation-des-donnees-de-sante-experimentee-au-chu-de-brest-avec-une-startup/.
[81] Johansson V., et al., (2022), "What ethical approaches are used by scientists when sharing health data? An interview study", *BMC Medical Ethics*, 23:41.

qui nécessitent des données réparties dans divers centres cliniques ou hôpitaux, mettant en évidence l'intérêt de mutualiser les données.

- Recommandation n°7 : Inciter les créateurs de PDS publiques à adopter des formats standards ouverts et des algorithmes *open source* pour mettre en qualité les données et traiter ultérieurement des flux de données, et aussi permettre des études multicentriques, afin de libérer le potentiel d'innovation de tous les réutilisateurs de données de santé.

Sur l'anonymisation :

- Recommandation n°8 : Développer la recherche sur les méthodes alternatives à l'anonymisation et à la pseudonymisation des données, notamment les techniques de chiffrement homomorphe, pour pouvoir mieux exploiter des données de santé.

II. SOUVERAINETE, AUTONOMIE ET VALORISATION DES PLATEFORMES DE DONNEES DE SANTE

1. Enjeux éthiques de la souveraineté des plateformes de données de santé

Dans un article de *La Croix* daté du 28 juin 2022[82], la journaliste Marion Durand rappelle les résultats d'un sondage Ifop[83] selon lequel 52 % des Français ne font confiance à aucun pays pour protéger leurs données personnelles, et seuls 10 % privilégieraient un acteur européen. Elle avance que la suspicion des Français a été nourrie par les différents usages des données personnelles à des fins commerciales ou politiques et les récents cas de fuites de données (en hausse de 19 % en 2020 selon le Forum international de cybersécurité), notamment de données médicales.

1.1 Une géopolitique des données de santé complexe

Les tensions que nous observons au niveau national, entre la nécessité et la volonté de protéger les données individuelles tout en les rendant accessibles pour le développement de la science et des projets d'intérêt commun se retrouvent au niveau international. En effet, chaque pays a intérêt à la fois à protéger et donc à limiter l'accès à ses données, et en même temps à bénéficier des données des autres et donc à partager ses propres données pour participer aux progrès de la connaissance. Ces tensions prennent des formes spécifiques au niveau géopolitique. D'abord, on observe une différence de modèle juridique entre le RGPD européen qui conçoit les données de santé comme un attribut de la personne et le modèle plus libéral états-unien qui les conçoit comme des biens susceptibles d'être commercialisés. Ceci peut mener à des situations complexes comme celle de l'Irlande qui cherche à attirer à soi les plateformes y compris américaines et qui, de ce fait, a du mal à appliquer les règlements européens en matière de respect de vie privée[84]. À l'inverse, il est important pour les acteurs d'un pays de pouvoir s'insérer, y compris avec ses données, dans les réseaux internationaux de recherche ou de développement pour participer aux avancées mondiales qu'ils peuvent engendrer[85].

Face à ces difficultés, le débat public en France s'est focalisé sur la notion d'une souveraineté des données à défendre.

1.2 Ambivalence de la notion de souveraineté

La souveraineté est une notion complexe qui provient du droit, de la philosophie et des sciences politiques. La conception classique et « fermée » de la souveraineté désigne un pouvoir exercé sur un territoire protégé par des frontières. Depuis l'époque moderne, elle se trouve associée et confrontée à la conception entrepreneuriale et « ouverte » de la souveraineté déterritorialisée d'acteurs économiques contrôlant des flux financiers et commerciaux.

---

82 Durand M., « L'altruisme des données, une utopie ? », *La Croix*, 28 juin 2022.

83 Sondage Ifop, *Les Français et la souveraineté numérique*, avril 2021.

84 Voir : https://www.lemonde.fr/pixels/article/2021/09/13/protection-des-donnees-l-irlande-maillon-faible-du-rgpd_6094434_4408996.html

85 Voir : https://www.udninternational.org/ *Undiagnosed diseases network international* est un exemple de réseau international de données génomique visant à mieux diagnostiquer les maladies très rares dans lequel les généticiens de tous pays sont admis, à condition de partager des cas cliniques.

Si la conception classique est associée à une idée de protection qui pourrait être unilatéralement imposée, la conception entrepreneuriale impose que la défense de la souveraineté soit négociée, souvent dans un cadre international multilatéral.

Cette distinction et cette ambivalence se retrouvent dans la notion de souveraineté numérique[86]. Il est intéressant de noter que le rapport Villani sur l'IA[87] ne mentionne jamais la « souveraineté numérique », mais l'englobe dans une problématique plus vaste de « souveraineté technologique et économique ».

On retrouve cette ambivalence en ce qui concerne les données de santé considérées à la fois comme un « trésor national » à protéger et un « bien commun » à partager, à une échelle européenne voire mondiale. Le système centralisé de la Sécurité sociale instauré par la loi du 26 janvier 2016 a eu comme effet bénéfique inattendu de faciliter la construction d'une base de données quasiment exhaustive sur la santé de la population française. Ce Système national des données de santé (SNDS[88]) représente aujourd'hui l'une des sources les plus complètes et riches au monde. Mais la question de savoir comment exploiter au mieux cette source d'informations implique de repenser le rôle de l'Etat dans la conduite des politiques d'innovation, et en particulier la défense de sa souveraineté dans la mesure où la santé est une mission régalienne de l'Etat, ce qui a conduit au projet du GIP-PDS[89]. Ces enjeux de souveraineté sur les données de santé se retrouvent aux échelles plus réduites d'une région, avec l'exemple du *Ouest Data Hub*[90] (annexe 4.3), ou d'un Centre hospitalier universitaire (CHU) qui veulent à la fois protéger et exploiter leurs propres données de santé. Les mêmes préoccupations se retrouvent à l'échelle de l'Union européenne qui, à travers la « *Gaia-X European Association for Data and Cloud* »[91], entend promouvoir les valeurs de protection des données, de transparence, de sécurité et de respect des droits sur les données.

1.3 Une vision libérale et entrepreneuriale pour une souveraineté conquérante

La considération d'une souveraineté portant d'abord sur l'exploitation des données a conduit à promouvoir une vision libérale et entrepreneuriale des plateformes de données de santé (PDS). Dans l'impulsion donnée par le rapport Villani sur l'Intelligence Artificielle, les données de santé sont d'abord envisagées comme devant permettre à la France de devenir un « leader mondial de la santé numérique » de deux manières complémentaires. Premièrement cette profusion de données mises à disposition devrait susciter une forte attractivité mondiale de la France où les entreprises existantes, françaises ou étrangères, développant des services numériques en santé et les chercheurs travaillant sur ce sujet pourraient trouver des données à partir desquelles travailler. Deuxièmement, un tel écosystème des données de santé devrait engendrer en France, du côté scientifique, des innovations de rupture et, du côté des entreprises, des « licornes », c'est-à-dire des start-up valorisées à plus de 1 milliard de dollars, censées permettre de rivaliser avec les GAFAM[92], ou à tout le moins avec les multinationales anglo-saxonnes. Ces licornes

---

[86] Ganascia J-G., Germain E., Kirchner C., (2018), *La souveraineté à l'ère du numérique Rester maîtres de nos choix et de nos valeurs*, CERNA, 36 p. http://cerna-ethics-allistene.org/digitalAssets/55/55708_AvisSouverainete-CERNA-2018.pdf

[87] Villani C., (2018), Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne, 235 p. [https://www.vie-publique.fr/sites/default/files/rapport/pdf/184000159.pdf].

[88] Voir annexe 4.1 et https://www.snds.gouv.fr/.

[89] Voir annexe 4.2 et https://www.health-data-hub.fr/

[90] Voir : https://www.chu-hugo.fr/accueil/projets/Ouest-DataHub/

[91] Voir : https://www.gaia-x.eu/

[92] GAFAM est l'acronyme des cinq grandes firmes américaines spécialisées dans le Web : Google, Apple, Facebook, Amazon et Microsoft.

constitueraient ainsi le bras armé – ou plutôt la corne défenderesse – de la France et de l'Europe dans la compétition économique internationale. La souveraineté, ici, signifie donc la capacité nationale à résister à la domination d'acteurs économiques étrangers (principalement américains) en développant notre propre puissance économique à l'échelle internationale.

En France, cette vision libérale s'accommode du rôle central joué par l'État dans la défense de cette souveraineté, ce qui se retrouve par exemple dans le rapport du député Eric Bothorel intitulé *Pour une nouvelle ère de la politique publique de la donnée*[93]. Conformément à ce qu'a pu écrire Mariana Mazzucato[94], l'État devrait faire les investissements initiaux, que les entreprises ne font pas parce qu'ils sont trop risqués, pour développer non pas une entreprise, mais tout un marché. Pour ce faire, le gouvernement a, d'une part, mis en place le GIP-PDS doté de presque 10 millions d'euros en 2020, dont le projet est de centraliser les données de santé et les rendre facilement accessibles aux chercheurs et aux start-ups et, d'autre part, établi, par l'intermédiaire de la Banque Public d'Investissement (Bpifrance), un plan de financement appelé DeepTech dont la filière la plus structurée, la santé, a injecté plus de 100 millions d'euros depuis 2019 dans le financement de start-ups, dont certaines s'appuient précisément sur les données offertes par le GIP-PDS. Ainsi, l'État a construit un « bac à sable », à savoir un espace où les investisseurs, les entrepreneurs et les chercheurs peuvent jouer avec les outils qui leur sont offerts.

Mais, certaines personnes ayant fondé l'association InterHop qui « promeut et développe l'utilisation des logiciels libres et *open-source* pour la santé », associés à d'autre acteurs militants du numérique, ont pointé une grave contradiction dans ce système. Alors que la France voulait construire la souveraineté française en se fondant sur des innovations de rupture et des licornes[95], le cahier des charges du GIP-PDS est tel que la solution retenue pour héberger les données de santé de manière centralisée est Microsoft Azure. Ce choix résulte du fait que Microsoft Azure semblait à ce moment-là capable de fournir plus rapidement que d'autres les fonctionnalités attendues et la sécurisation exigée pour des données aussi sensibles. Ce choix du GIP-PDS de la technologie Microsoft Azure a été vivement dénoncé par InterHop[96] qui a saisi, avec d'autres requérants, le Conseil d'État. L'ordonnance du juge des référés n°444937 du 13 octobre 2020 prend acte de l'intention annoncée par le Gouvernement à l'occasion de cette action en justice d'adopter, dans les délais les plus brefs possibles, des mesures propres à éliminer tout risque, telles que le choix d'un nouveau sous-traitant, ou le recours à un accord de licence, suggéré par la Commission nationale de l'informatique et des libertés (CNIL). Le juge des référés a aussi prescrit au GIP-PDS de veiller à la mise en œuvre par Microsoft des mesures techniques et organisationnelles appropriées pour garantir au mieux la protection des droits des personnes concernées. Il a enfin jugé qu'il existait un intérêt public important à permettre la poursuite de l'utilisation des données de santé pour les besoins de la gestion de l'urgence sanitaire et de l'amélioration des connaissances sur le SARS-CoV-2 et, à cette fin, de permettre le recours aux moyens techniques, sans équivalent à ce jour, dont dispose

[93] Bothorel E., *Pour une nouvelle ère de la politique publique de la donnée*, rapport de la mission parlementaire présidée par Éric Bothorel remis au Premier ministre le 23 décembre 2020, 216 p.

[94] Mazzucato, M., (2013), *The Entrepreneurial State: Debunking Public vs. Private Sector Myths*. London: Anthem Press, 288 p.

[95] Voir : https://www.elysee.fr/emmanuel-macron/2017/09/20/discours-d-emmanuel-macron-devant-la-72e-assemblee-generale-des-nations-unies.

[96] Voir : https://interhop.org/2020/12/14/stophealthdatahub-donnees-de-sante-en-otage-chez-microsoft.

le GIP-PDS par le biais du contrat passé avec Microsoft. Une solution d'hébergement européenne, d'abord envisagée en 2022, a été finalement repoussée à 2025[97].

1.4 Une vision régulatrice pour une souveraineté protectrice

La critique portée par certains gestionnaires de données, reprise et complétée par de nombreux acteurs dont la CNIL et le Conseil d'État, a concouru à faire émerger une deuxième notion de souveraineté, que l'on peut appeler légaliste et protectrice. Elle est défendue par exemple par la sénatrice Catherine Morin-Desailly et par la commission parlementaire trans-partisane créée en juin 2020, sur le thème « Bâtir une souveraineté numérique française et européenne », dont le député Philippe Latombe est rapporteur. Selon cette perspective, il ne s'agit pas de remettre en cause le projet de faire fructifier le trésor des données de santé, ni de remettre en cause la création d'un GIP-PDS, mais de protéger les données elles-mêmes (et non les marchés) contre les appropriations extérieures à l'Union Européenne. Ces données sont les attributs des personnes résidant en France, et ne doivent pas être transférées sur des serveurs étrangers non européens, au risque que toutes les protections offertes par le RGPD, qui ne s'applique qu'en Europe, soient inutiles. Ici, la souveraineté englobe la capacité à protéger les données personnelles de la population contre des usages illégaux, au sens de la loi française d'abord, européenne ensuite.

À nouveau, pour promouvoir cette souveraineté, l'État joue un rôle crucial, mais très différent du précédent : certes, il doit créer un GIP-PDS, en particulier pour faire avancer la science, mais il doit en même temps promouvoir la création d'un « *Cloud* Souverain », c'est-à-dire dont les infrastructures sont implantées physiquement en Europe afin que la législation européenne s'applique. Il doit mettre en place une gouvernance du GIP-PDS plus attentive aux critiques émises par différents acteurs nationaux de l'écosystème numérique, en particulier la Caisse nationale d'assurance maladie et de travailleurs salariés (CNAMTS) qui, pour l'instant, héberge une partie importante des données via le SNDS et bénéficie à ce titre d'une expérience qu'elle propose de partager, tout en ne cherchant pas nécessairement à centraliser toutes les données. Enfin, l'Etat doit établir les lois grâce auxquelles ces données sont protégées et les usages illicites sanctionnés.

1.5 Une vision européenne fondée sur la notion d'autonomie stratégique

1.5.1 L'association Gaia X

C'est hors du cadre institutionnel de l'Union européenne, mais à l'intérieur des frontières européennes, qu'à l'initiative de la France et de l'Allemagne, s'est créée en 2021 l'association Gaia-X qui comprend actuellement plus de 350 entreprises et organisations, dont les GAFAM [98].

Son objectif n'est pas de créer un géant numérique européen destiné à assurer une réelle souveraineté économique européenne, mais de favoriser la constitution en Europe d'une fédération logicielle en réseau susceptible de connecter des fournisseurs de services *cloud* et des propriétaires de données dans un environnement de confiance et de stimuler la création de nouveaux espaces de données communs dans différents domaines dont la

[97] Déclaration de la directrice du GIP-PDS en septembre 2022 au média Tic Pharma. Voir : https://www.ticpharma.com/story/2044/stephanie-combes-devoile-le-programme-de-rentree-du-health-data-hub.html

[98] Mais seuls des membres européens peuvent siéger au conseil d'administration : https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html

santé, en respectant des règles strictes de portabilité, d'interopérabilité, d'autodétermination sur les données et de sécurité. L'association Gaia-X définit le concept de « souveraineté sur les données » (*data sovereignty*) comme « la capacité, pour les participants du marché des données, de pouvoir s'autodéterminer en ce qui concerne l'échange et le partage de données » et « de prendre des décisions éclairées sur les services qui adhèrent à des spécifications techniques spécifiques et à des réglementations européennes ou nationales de leur choix »[99]. Gaia-X permettrait ainsi à une entreprise située dans un État membre de l'Union européenne de trouver une solution d'hébergement *cloud* dans un autre État membre, de recourir à de la puissance de calcul dans un troisième État, en faisant appel à une interface de gestion des données qui pourrait être située dans un quatrième État, tout en respectant les normes de sécurité et de protection des données en vigueur en Europe[100].

Pour atteindre ces objectifs, l'association décernera aux opérateurs des labels de niveaux différents selon notamment la garantie qu'ils offrent en matière de localisation du traitement des données et des services en Europe et leur immunité aux lois non-européennes. Le label le plus exigeant est prévu pour les données de santé.

Même si l'association se réfère dans ses présentations à la notion de souveraineté numérique, l'inclusion d'acteurs non européens dans un projet à vocation européenne peut interroger sur la pertinence du choix de ce terme. L'association semble plus proche de la position française, telle qu'énoncée par Henri Verdier, ambassadeur pour le numérique : « Beaucoup voient la question de la souveraineté comme une question d'hégémonie. Or, nous la voyons comme une question d'autonomie stratégique »[101]. Cette dernière expression est apparue dans le vocabulaire de l'Union européenne en 2013 au sujet de la défense[102] puis a été reprise en 2020[103] de manière plus large et appliquée notamment au numérique.

1.5.2 L'espace européen des données de santé

Sur le plan institutionnel européen, s'agissant des données de santé, le projet de règlement pour un Espace européen des données de santé publié par la Commission européenne en mai 2022[104] est l'exemple d'un outil numérique européen fondé sur la notion d'autonomie stratégique. Le comité de protection des données de l'Union européenne a, dans son avis du 12 juillet 2022 sur ce projet, insisté sur le fait que les données devraient être hébergées en Europe[105].

1.6 Tensions éthiques entre visions de la souveraineté et de l'autonomie des PDS

Il y a donc au moins trois conceptions de la souveraineté. Les tenants d'une vision libérale de la souveraineté reprochent à ceux qui en défendent une vision régulatrice de ralentir la marche des entreprises françaises, et, ce faisant, de leur faire perdre leur place dans la compétition internationale ; ils craignent que les réglementations étouffent les start-ups

[99] Voir : *La revue européenne des médias et du numérique*, n°54 bis- 55 automne 2020 [https://la-rem.eu/2021/01/un-chiffre-ou-deux-n54bis-55-automne-2020?action=genpdf&id=15339].

[100] Voir : *La revue européenne des médias et du numérique*, n°54 bis- 55 automne 2020 [https://la-rem.eu/2021/01/un-chiffre-ou-deux-n54bis-55-automne-2020?action=genpdf&id=15339].

[101] Renaissance numérique, (2022), « Rapport sur la souveraineté technologique européenne », p. 16. [https://www.renaissancenumerique.org/publications/la-souverainete-technologique-europeenne/]

[102] Conseil européen des 19 et 20 décembre 2013

[103] Conseil européen d'octobre 2020

[104] Commission européenne, "Proposal for a regulation - The European Health Data", 3 mai 2022 [Spacehttps://health.ec.europa.eu/publications/proposal-regulation-european-health-data-space_fr].

[105] Voir : https://edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion-032022-proposal_fr

sans contrôler les multinationales américaines capables, grâce à leur puissance, de contourner les règles. *A contrario*, les régulateurs reprochent aux libéraux de se méprendre sur le « ruissellement » qu'ils attendent de la création des licornes au motif que ces dernières, si elles accumulent parfois d'énormes capitalisations, produisent une valeur ajoutée minime, et laissent toujours courir le risque que ces mêmes capitalisations soient rapidement revendues à des sociétés concurrentes étrangères et non-européennes en particulier.

La conception de l'autonomie stratégique telle qu'elle est mise en œuvre par Gaia-X privilégie la décentralisation et la fédération. Elle évite de créer de nouveaux acteurs monolithiques et permet de travailler à partir de l'existant. Mais est-elle suffisamment exigeante et ne prend-elle pas des risques en admettant des entreprises étrangères à l'Union européenne au sein de l'association ? Inspirée par le droit de la *compliance*[106], l'autonomie stratégique selon Gaia-X, qui implique une internalisation de principes par l'entreprise, exclut un régulateur extérieur : l'architecture Gaia-X s'autocontrôle. Or, un opérateur extra-européen, ayant créé une société ayant la nationalité d'un État membre, ne va-t-il pas pouvoir bénéficier du label le plus exigeant, s'il réussit à s'intégrer dans une structure de droit européen qui respecte les critères, alors qu'on pourra toujours s'interroger sur son indépendance à l'égard de la législation de l'État étranger dont il relève ?

D'où l'importance, avant la création et la certification d'une PDS, d'adopter une démarche pluridisciplinaire, avec des experts en santé, en informatique, en droit et en sciences humaines et sociales pour réfléchir aux enjeux de souveraineté de cette plateforme et anticiper les risques d'atteinte à sa souveraineté (voir Recommandation n°9).

Mais ce n'est pas le devenir des entreprises qui est l'enjeu principal. La valeur éthique que l'on peut attribuer à ces visions de la souveraineté appliquées aux PDS ne peut se rapporter qu'à leurs rapports au bien commun et aux principes spécifiques de la bioéthique : bienfaisance et non-malfaisance, promotion et respect de l'autonomie, justice et équité, auxquels s'ajoutent les principes d'explicabilité et de transparence, spécifiques de l'éthique de l'IA (voir §2 de l'introduction). À cet égard on attend d'abord de l'État de veiller simultanément, via les PDS, à l'amélioration de la santé publique, à sa justice et son équité, et à la protection des données de santé de chaque citoyen.

Une première tension éthique entre les différentes visions de la souveraineté concerne le progrès en matière de recherche médicale et de soins, et donc le principe de **bienfaisance**, ce qui suppose une facilitation des accès aux données de santé et donc de leurs échanges, et tendrait à privilégier l'efficacité et la performance rapide selon la vision libérale et entrepreneuriale de la souveraineté qui a conduit en particulier au choix de Microsoft Azure pour le GIP-PDS.

Une deuxième tension éthique porte sur le risque de prise de contrôle par des sociétés non-européennes sur ce bien commun national que constituent nos bases de données de santé, que ce soit par une prise de contrôle des PDS ou par une exploitation de leurs données. Ceci pourrait entraîner de graves atteintes à la **justice** et à l'**équité** qui prévalent actuellement dans le système de santé et de protection sociale français si, malgré la réglementation actuelle en France ou en Europe, telle ou telle de ces sociétés en venait à profiler les personnes par leurs données de santé et à leur proposer ensuite des contrats d'assurance différentiés en fonction de ce profilage (voir Recommandation n°10).

[106] Frison-Roche M-A., (2019), « L'apport du droit de la compliance à la gouvernance d'Internet », Rapport commandé par Monsieur le Ministre en charge du Numérique, 134 p.

Une troisième tension éthique porte sur les principes d'**explicabilité** et de **transparence** qui pourraient être mis en péril si l'on externalise des fonctions par l'incapacité de l'État français ou de l'Europe d'exercer un droit extraterritorial sur un partenaire étranger comme c'est le cas actuellement, hormis pour le Règlement général sur la protection des données (RGPD), contrairement à ce que peuvent revendiquer les États-Unis. Ce risque, comme la menace de mainmise par une société privée, tendrait à privilégier la vision légaliste et protectrice de la souveraineté. Mais quel que soit le modèle de souveraineté retenu, il est essentiel de prévoir des clauses de **réversibilité** sur le rôle joué par telle ou telle entreprise pour s'assurer que i) les données sont stockées sans recours à un format propriétaire et pourront être facilement transférées aux utilisateurs, en particulier aux chercheurs, à des coûts minimaux, ii) les algorithmes seront développés avec des technologies de référence, indépendantes de la solution d'hébergement, iii) les éléments de la plateforme seront en partie réutilisables (architecture, scripts de déploiement, etc.)[107] (voir Recommandation n°11).

Une quatrième tension éthique concerne le principe de **non-malfaisance** devant les risques d'exploitation des données de santé pour développer des médications nocives et ciblées sur des catégories de personnes vulnérables, au risque de cyberattaques sur les PDS. Face à cela, ce sont l'architecture et l'infrastructure des PDS qui sont en cause sans que cela conduise *a priori* à privilégier une vision particulière de la souveraineté numérique. Un point de vigilance particulier s'impose lors des opérations de maintenance des systèmes informatiques servant les PDS et impliquant des acteurs non européens. Ces contextes particuliers peuvent nécessiter des stockages momentanés de données et accroissent leur vulnérabilité à des cyberattaques et à des accès non autorisés, voire à des exportations de données de santé (voir Recommandation n°12).

2. Formes de valorisation des plateformes de données de santé

Comme nous l'avons précisé au § I.1, les données ne sont pas cessibles, mais il est possible de vendre un droit d'accès à ces dernières. Or, l'ensemble des PDS rencontrent certaines difficultés pour établir la grille tarifaire de leur utilisation. Ces obstacles peuvent s'expliquer parce que le service qui est monnayé, à savoir l'accès à d'importantes bases de données de santé, est émergent et se trouve en tension entre deux « formes de mise en valeur » des données. Ce concept, qui a été mis en évidence par Boltanski et Esquerré[108], désigne des formes conventionnelles de valorisation, propres aux sociétés dans lesquelles on les trouve, mais suffisamment stabilisées pour constituer « une ressource collective à laquelle les agents peuvent se référer quand ils doivent s'orienter dans le monde des objets [pour leur attribuer une valeur] ».

Or les données ne sont pas des objets à proprement parler. Ce sont des biens immatériels, en tant qu'informations incessibles d'une part, et « non rivaux », d'autre part, c'est-à-dire qu'ils peuvent être consommés simultanément par plusieurs personnes sans entraîner de perte directe du bien (plusieurs personnes peuvent travailler simultanément sur une même base de données sans la dégrader), ce qui est impossible avec des objets matériels. Pourtant, l'intuition qui est à la source de la notion des « modes de valorisations » reste utile à notre réflexion. En effet, les bases de données de santé auxquelles les plateformes vendent l'accès peuvent être envisagées selon deux formes de mise en valeur différentes.

[107] Voir : https://www.health-data-hub.fr/sites/default/files/2021-05/Etude%20de%20r%C3%A9versibilit%C3%A9%20de%20la%20plateforme%20technologique%20%E2%80%93%20Novembre%202019.pdf;

[108] Boltanski, Luc, et Arnaud Esquerre. *Enrichissement: une critique de la marchandise*. NRF essais. [Paris]: Gallimard, 2017.

2.1 Valorisation sur la base du coût de création et de maintenance

Premièrement, depuis une première perspective dite « standard » (pour reprendre le terme de Boltanski et Esquerré), les bases de données sont envisagées depuis la perspective du travail et des investissements nécessaires à leur constitution et à leur maintenance. Elles reposent sur un travail de collecte, de stockage et de mise en œuvre de briques technologiques éventuellement fournies par des partenaires de la PDS. La question est donc de parvenir à estimer le coût de ces services, par exemple au moyen de comptabilité analytique interne aux organisations, éventuellement d'y ajouter une part de valeur ajoutée pour estimer leur valeur et leur prix d'accès. Concrètement, on a pu observer que la rémunération des diverses fonctions et services d'une PDS repose sur divers mécanismes de tarification : prix d'entrée, droits d'auteur sur le produit final, prix libératoire, coopération gratuite pour créer des brevets.

2.2 Valorisation sur la base des bénéfices futurs espérés

Mais, deuxièmement, la base de données peut être envisagée depuis une autre perspective de mise en valeur, que Boltanski et Esquerré appellent la forme « actif » selon laquelle les bases de données mises en forme sont envisagées comme des parts de capital qui permettent de produire une valeur future très supérieure à celle d'aujourd'hui grâce à des mécanismes financiers de revente de participation dans une entreprise. Dans ce cas, l'espérance de valorisation future entre dans le calcul du prix. Pour reprendre les caractéristiques identifiées par un *benchmark* des modes de valorisation des bases de données de santé, commandé par l'alliance Aviesan[109], elles ont une « capacité à faire émerger de la valeur », elles comportent un risque – et donc une opportunité – associé au partenariat entre la plateforme détenant les données et l'acteur qui les utilise, ou encore sont l'objet d'accords qui incitent plus ou moins les personnes qui s'engagent dans le projet à s'investir. La valeur attribuée aux données est déterminée par des mécanismes comme la création d'une *joint-venture* entre la plateforme et l'entreprise développant un projet, ou encore le partage d'un pourcentage du chiffre d'affaires de cette dernière.

2.3 Débat européen et français entre les deux formes de valorisation

Ces deux modes de valorisation ont tous deux été soutenus par des institutions puissantes qui les légitimaient tout autant l'un que l'autre. Pour le dire vite, le premier modèle est porté par le système de santé français centré sur les CHU et les institutions publiques de recherche, quand le second est porté par des acteurs comme la Banque Publique d'Investissement et son plan *Deeptech*[110] et par de nombreuses start-up.

2.3.1 Valorisation selon le règlement européen sur la gouvernance des données

Le règlement européen du 30 mai 2022 sur la gouvernance des données (*Data Governance Act*[111]) opte pour la première branche de l'alternative pour faciliter le droit d'accès aux données, en retenant des redevances faibles. Ce texte comprend des dispositions sur le cadre général de mise à disposition de différents types de données, dont les données personnelles, par les organismes du secteur public intégrant notamment l'État, les autorités régionales ou locales ou les organismes de droit public. Diverses garanties sont prévues dont notamment l'anonymisation. Le but étant de favoriser l'accès

[109] La perspective de la forme « actif » des bases de données de santé est celle qui est adoptée par l'alliance Aviesan dans son *benchmark* de différents « modèles de valorisation des données de santé ». Voir : https://cvt.aviesan.fr/outils/enjeux-lies-aux-donnees-de-sante/.

[110] Voir : https://www.bpifrance.fr/nos-actualites/plan-deeptech-3-chiffres-2-ans-un-seul.

[111] Voir : https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R0868. Publié au Journal Officiel de l'Union Européenne le 23 juin 2022, le DGA entrera en vigueur en septembre 2023.

à ces données, le niveau des redevances versées pour la réutilisation de ces données sera uniquement calculé sur la base des « coûts liés à la conduite de la procédure de réutilisation des catégories de données » mises à disposition (article 6. 5). Ainsi, la valeur réelle d'accès et d'utilisation de telles bases de données ne sera pas prise en compte, puisque les coûts retenus seront essentiellement ceux de la fourniture et de la diffusion des données, de l'anonymisation ou autres formes de préparation des données, de la maintenance de l'environnement de traitement sécurisé.

2.3.2 Quel financement pour les plateformes de données de santé en France ?

Ainsi, partant du constat que les financements actuels des plateformes, qui reposent le plus souvent sur des programmes d'investissement d'avenir non récurrents, sont insuffisants et trop incertains, la Première ministre a confié au Comité stratégique des données de santé[112] la mission d'analyser et de questionner la pertinence de nos outils actuels de régulation et de financement des produits de santé pour formuler ses recommandations d'ici l'été 2023.[113] Il s'agit donc de financer les PDS de façon pérenne à un niveau suffisant pour qu'elles n'aient pas besoin de revenus supplémentaires. Et dans le même temps, d'établir une grille tarifaire pour certains acteurs, dans le cas de missions d'intérêt public, qui soit abordable pour qu'ils y aient accès relativement aisément. Ces deux orientations reviennent à envisager les données de santé depuis le mode de valorisation standard (voir Recommandation n°13).

2.3.3 Le projet de règlement européen sur les données

La Commission européenne a présenté le 23 février 2022 une proposition législative, le Règlement sur les données (*Data Act ou* DA), dont l'objectif[114] est d'assurer une meilleure répartition de la valeur issue de l'utilisation des données personnelles et non personnelles entre les acteurs de l'économie de la donnée, notamment liées à l'utilisation des objets connectés et au développement de l'Internet des objets (*IoT, Internet of Things*). Ce point concerne les données de santé recueillies par tous les dispositifs médicaux connectés et qui peuvent alimenter des plateformes de données de santé. Cet objectif du DA va bien dans le sens de notre recommandation sur le partage de la valeur (voir Recommandation n°14).

2.4 Vigilance concernant les conflits d'intérêts

Dans la mesure où les formes de valorisation des plateformes de données de santé peuvent faire intervenir des intérêts privés à côté de l'intérêt public elles n'échappent pas au risque de conflits d'intérêts. [115] Ceci appelle donc à une vigilance dans les nominations aux postes de responsabilité et les clauses de non-concurrence pour les dirigeants de PDS publiques (voir Recommandation n°15).

---

[112] Créé par un arrêté du 29 juin 2021. Voir : https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000043850566

[113] Voir : https://www.gouvernement.fr/communique/mecanismes-de-regulation-et-de-financement-des-produits-de-sante

[114] CNIL, « Stratégie européenne pour la donnée : la CNIL et ses homologues se prononcent sur le Data Governance Act et le Data Act » [https://www.cnil.fr/fr/strategie-europeenne-pour-la-donnee-la-cnil-et-ses-homologues-se-prononcent-sur-le-data-governance].

[115] Voir : https://www.lemonde.fr/planete/article/2019/12/24/donnees-de-sante-conflit-d-interets-au-c-ur-de-la-nouvelle-plate-forme_6023918_3244.html

3. Recommandations

- **Recommandation n°9** : Exiger pour la création et la certification d'une PDS une démarche pluridisciplinaire, avec des experts en santé, en informatique, en droit et en sciences humaines et sociales pour anticiper les risques d'atteinte à la souveraineté.
- **Recommandation n°10** : Dans les contrats de partenariat internationaux impliquant des données de santé, veiller à ce qu'il y ait des clauses garantissant que les acteurs non européens respectent les principes fondamentaux du RGPD, du règlement sur la gouvernance des données, et du futur règlement européen sur les données, pour protéger les données sensibles.
- **Recommandation n°11** : Prévoir systématiquement des clauses spécifiques de transparence, d'explicabilité et de réversibilité, pour les entreprises, notamment extra-européennes, permettant en particulier des transferts de données à des coûts minimaux.
- **Recommandation n°12** : Veiller aux conditions d'accès aux données de santé et d'exportation temporaire lors de la maintenance des systèmes informatiques servant les PDS impliquant des acteurs non européens.

Sur la valorisation des données

- **Recommandation n°13** : Encourager un financement des PDS sur la base de leurs coûts d'investissement et de fonctionnement, et une tarification adaptée à différents usagers, en particulier pour la recherche scientifique d'intérêt public.
- **Recommandation n°14** : Inciter les entreprises qui rencontrent le succès financier en partie grâce à des jeux de données fournis par des plateformes publiques de données de santé, à partager une partie de leurs bénéfices avec ces dernières, par la signature volontaire d'une charte engageant leur réputation.
- **Recommandation n°15** : Veiller à maintenir une indépendance forte entre les équipes de direction des plateformes publiques de données de santé et celles des entreprises utilisatrices afin de prévenir les conflits d'intérêts.

Encadré n° 4 : enjeu de souveraineté sur les PDS de radiologie

Après les cliniques privées, les EHPAD, et plus récemment les plateformes de biologie, largement acquis par des groupes financiers relevant de fonds d'investissements internationaux, les radiologues du secteur libéral, avec leurs plateaux techniques d'imagerie médicale font l'objet de propositions d'acquisition massive par des investisseurs, dont l'offre financière attrayante masque bien des incertitudes et des risques.

Le premier de ces risques concerne **l'absence de transparence du montage** de nombre de sociétés qui se portent acquéreurs des plateformes professionnelles avec trois niveaux de structure : les SEL (Sociétés d'Exercice Libéral), les SELAS, (Sociétés d'Exercice Libéral par Actions Simplifiées) et les Holdings Financières partenaires dans les SELAS, avec pour conséquences que les investisseurs étrangers n'apparaissent jamais directement dans le capital des Sociétés auxquelles les professionnels de santé vont se lier par contrat.

Ce montage ouvre la voie à une **double opacité des contrats proposés**. Une première opacité résulte du fait que les Conseils Départementaux de l'Ordre des Médecins ne donnent leur avis que sur les statuts des SEL, et aucunement sur les contrats connexes et

complexes qui sont signés par ailleurs et qui ne leur sont pas communiqués alors qu'ils le devraient. Une deuxième opacité apparaît pour les professionnels qui n'ont pas la maîtrise de la gouvernance, de la gestion et du contrôle des droits financiers échappant aux radiologues associés exerçant dans ces SELAS.

Ces contrats, à multiples étages sur le fond et la forme, induisent **en premier lieu** une **dérèglementation des professions réglementées** avec des risques avérés de perte d'autonomie décisionnelle et d'orientation d'activités vers des examens rentables, simples et modélisés, ainsi qu'une possible atteinte au libre choix des patients par la signature de clauses d'exclusivité entre le groupe et certaines cliniques privées ou plateformes de téléradiologie (y compris à l'étranger). Cela conduit à un risque évident de pratiques compromettant l'indépendance des professionnels, pourtant garantie par l'article R. 4127-5 du Code de la santé publique, ainsi que d'atteinte majeure au respect de la liberté de choix du patient : ainsi les patients pourraient ne plus avoir accès à un spécialiste qu'ils ont choisi ou que leur conseille leur médecin.

En second lieu, il convient d'alerter sur la **nécessité de clarifier la propriété des données massives d'imagerie de patients possiblement interprétées, stockées et exploitées à l'étranger** (dans le cadre de réseaux de téléimagerie) et dès lors échappant à tout contrôle, malgré le financement par les fonds publics de l'Assurance Maladie de ces examens, et compte tenu de la sensibilité des données diagnostiques et thérapeutiques personnelles qu'elles contiennent.

III. CONSENTEMENT AU PARTAGE DES DONNEES ET PARTICIPATION CITOYENNE A L'ELABORATION ET A LA GOUVERNANCE DES PLATEFORMES DE DONNEES DE SANTE

1. Diversité des formes de consentement

Le traitement des données personnelles de santé est en général interdit, néanmoins le RGPD prévoit dans son article 9 des exceptions notamment « lorsque la personne concernée a donné son **consentement** explicite au traitement de ces données à caractère personnel pour une ou plusieurs finalités spécifiques » [116].

1.1 Consentement libre éclairé et spécifique

La définition du consentement est donnée dans l'article 4(11) et les modalités du recueil du consentement sont précisées dans l'article 7 du RGPD[117]. Le consentement doit être **libre** (la personne doit pouvoir donner son accord sans contrainte), **spécifique** (une ou plusieurs finalités doivent être indiquées, et le consentement doit être donné séparément pour chaque finalité), **éclairé** (la personne comprend le traitement qui sera fait de ses données et des garanties doivent être données que la personne est consciente des implications de son consentement) et **non ambigu** (la personne doit avoir explicitement donné son consentement et le responsable de traitement doit pouvoir le prouver). Sur ce dernier point, ce consentement est clairement un mécanisme **opt-in** (tant que les personnes ne disent pas explicitement oui, il faut considérer que c'est non), puisque le silence, des cases pré-cochées ou l'inactivité ne peuvent être considérés comme un consentement[118]. Enfin, la personne doit pouvoir **retirer** son consentement quand elle veut, et le processus pour le faire doit être facile.

Si les caractéristiques du consentement sont aisées à comprendre, la mise en œuvre des modalités du recueil de ce consentement est moins évidente. La CNIL donne quelques indications pour guider les responsables de traitements[119] et le CCNE a discuté récemment l'évolution des enjeux éthiques relatifs au consentement dans le soin dans son avis 136 (voir encadré n°5).

Encadré n° 5 : Avis n°136 du CCNE

L'avis n°136 du CCNE sur « L'évolution des enjeux éthiques relatifs au consentement dans le soin »[120] s'est penché sur les questions de l'effectivité du recueil du consentement et la complexification des enjeux éthiques liés du fait du développement de nouvelles techniques médicales. Pour saisir les enjeux en cours, le CCNE s'est attaché à considérer le consentement, acte de soin à part entière, comme un processus évolutif et dynamique

[116] Voir : https://www.cnil.fr/fr/reglement-europeen-protection-donnees/chapitre2#Article9.

[117] Voir : https://www.cnil.fr/fr/reglement-europeen-protection-donnees/chapitre2#Article7; Voir aussi : Considérants 32 (Conditions for consent), 33 (Consent to certain areas of scientific research), 42 (Burden of proof and requirements for consent), 43 (Freely given consent) du RGPD.

[118] Considérant 32: "Silence, pre-ticked boxes or inactivity should not therefore constitute consent".

[119] Conformité RGPD : comment recueillir le consentement des personnes ? CNIL, 2018, https://www.cnil.fr/fr/les-bases-legales/consentement.

[120] CCNE, avis n° 136 du 15 avril 2021, *L'évolution des enjeux éthiques relatifs au consentement dans le soin*, 51 p.

qui « ne se donne pas une fois pour toutes, mais s'élabore et peut évoluer dans le cadre d'une relation fondée sur une confiance réciproque »[121] incluant de possibles rétractations. Cet avis préconise notamment d'accroître le rôle de la personne de confiance, pour une approche plus respectueuse de la volonté des personnes vulnérables (ayant des difficultés à exprimer leurs volontés ou n'étant pas en mesure de décider pour elles-mêmes). En complémentarité du renforcement du rôle de la personne de confiance, le numérique et ses outils d'informations apparaissent comme des supports privilégiés pour aider à l'expression et à la mémoire du processus de consentement, en particulier en matière de traçabilité de l'information et du cheminement de la personne, selon le CCNE.

L'exigence de cette forme de consentement en matière de soins répond à plusieurs impératifs éthiques bien connus en bioéthique : respect du droit à l'auto-détermination (autonomie individuelle), bienfaisance, justice, et surtout respect de la personne et de sa dignité[122]. Dans le cas des plateformes de données de santé (PDS), il s'agit du respect de l'autonomie du patient, dans sa capacité à se déterminer concernant l'usage qui est fait de ses données.

La difficulté du consentement éclairé spécifique est que le consentement est accordé pour une finalité précise et une durée précise, ce qui va à l'encontre de la philosophie des PDS dont la fonction est de permettre des usages ultérieurs des données et non forcément prévus. Comment alors consentir ? À côté de ce modèle traditionnel de consentement, d'autres formes ont émergé pour les biobanques étrangères et la recherche sur les données de santé[123].

1.2 Autres formes de consentement

Par opposition au consentement spécifique, le **consentement général** (*broad consent*) demande aux individus de consentir à de multiples études à venir, sans connaître les spécificités de ces études au moment de la demande. L'information sur les buts, les risques et les bénéfices éventuels est donnée de manière globale. Les individus ont moins de contrôle sur les données, puisqu'il n'y a pas d'échanges réguliers, d'où une moindre possibilité de se rétracter s'ils ont oublié qu'ils ont donné leur consentement ou si les conditions d'utilisation des données changent[124].

C'est ce modèle de consentement qu'utilise la plateforme de données de santé du NIH (*National Health Institutes*), pour son programme de recherche *All of Us Research Program*[125] qui vise à faciliter le développement de la médecine de précision, à travers la mise en place d'une cohorte de recherche large, bien caractérisée, tout en imposant des règles d'utilisations très précises, en demandant par exemple aux chercheurs et institutions de fournir un plan de partage des données génomiques[126]. C'est aussi le cas pour le projet britannique *100,000 Genomes project* du NHS (*National Health Service*) lancé en 2018, qui a pour but de constituer une base de séquences des génomes d'environ 85 000 patients du NHS atteints d'une maladie rare ou d'un cancer, ainsi que

[121] CCNE, avis n° 136, *op.cit.* p.4.
[122] CCNE, avis n° 136, *op.cit*.
[123] Wiertz S., Boldt J., (2022), "Evaluating models of consent in changing health research environments", *Med Health Care Philos*, Jun;25(2):269-280.
[124] Mikkelsen RB, Gjerris M, Waldemar G, Sandøe P. : "Broad consent for biobanks is best - provided it is also deep". *BMC Med Ethics*. 2019 Oct 15, 20(1):71.
[125] Voir : https://allofus.nih.gov/about/protocol/all-us-consent-process
[126] Voir : https://sharing.nih.gov/genomic-data-sharing-policy/about-genomic-data-sharing/gds-policy-overview

de leur famille, pour mettre en relation ces maladies avec les gènes susceptibles de jouer un rôle dans leur apparition et leur développement.

Le modèle de **consentement dynamique** (*dynamic consent*)[127], quant à lui, n'est pas vraiment une forme de consentement mais repose sur une plateforme de communication en ligne personnalisée qui facilite le processus de consentement entre les chercheurs et les participants. Il permet une communication bidirectionnelle et peut renforcer le droit des participants à la recherche de faire des choix autonomes concernant leur participation à la recherche, d'améliorer leur compréhension du processus de consentement et de favoriser leur engagement dans l'entreprise de recherche. Il permet également de stocker les consentements. Ce modèle est à la base du consentement dynamique spécifique et du méta-consentement et présuppose chez les patients l'accessibilité à des plateformes numériques et une certaine « littératie numérique ». Il répond aux préconisations du CCNE dans son avis n°136, dans lequel le consentement est décrit comme un processus évolutif et dynamique, ce qui amène le comité d'éthique à préconiser d'utiliser l'outil numérique dans le cadre du recueil et de la mémoire du consentement.

Dans le modèle du **consentement dynamique spécifique** (*dynamic specific consent*)[128], les informations sont données aux personnes dans le format de leur choix, correspondant à leur niveau d'éducation et à leur intérêt. Les participants aux projets sont tenus régulièrement informés des modifications. Dans ce modèle, on peut envisager que des QCM permettent de s'assurer que les participants ont bien compris l'information qui leur était donnée. Le projet de plateforme génomique, *Promise for Engaging Everyone Responsibility* (PEER)[129], développé par la *Genetic Allianc*e demande le consentement dynamique spécifique et les participants aux recherches peuvent indiquer via une matrice les types d'accès qu'ils approuvent et ceux qu'ils n'approuvent pas.

À la différence du consentement dynamique spécifique qui propose le même formulaire de consentement à tous les participants, dans le modèle du méta-consentement (*meta consent)*[130], les personnes peuvent donner un consentement spécifique ou général, selon leurs préférences personnelles. Des catégories leur sont suggérées, pour permettre différentes options de consentements selon les domaines et les formes des projets. Les préférences des utilisateurs sont gérées par la plateforme et peuvent être changées à tout moment. Les chercheurs peuvent contacter les participants par requête sur la plateforme. Ce modèle de méta-consentement[131] est utilisé par la plateforme de données de santé du GRIIS[132] développée au Canada : « Au lieu de consentir à un projet à la fois, les patients et les personnes en général consentiraient à ce que plusieurs projets ayant des caractéristiques similaires puissent accéder à leurs données de santé ». Le principal défi de ce modèle est la détermination des catégories de projets, qui pourrait néanmoins bénéficier de la participation des patients.

1.3 Données de santé *post-mortem*

Alors que le RGPD précise ne pas s'appliquer aux données des personnes décédées, la loi du 6 octobre 2016 pour une République numérique a établi un cadre juridique régissant

[127] Budin-Ljøsne I. et al., (2017), "Dynamic Consent: a potential solution to some of the challenges of modern biomedical research", *BMC Med. Ethics*, 18, 4.

[128] *Ibidem*.

[129] Voir : https://geneticalliance.org/registries/promise-for-engaging-everyone-responsibly.

[130] Ploug T., Holm S., (2016), "Meta Consent – A Flexible Solution to the Problem of Secondary Use of Health Data", *Bioethics*, 30 (9), pp. 721 – 732.

[131] Voir : https://griis.ca/recherche/claret/.

[132] Cumyn A. et al., (2021), "Meta-consent for the secondary use of health data within a learning health system: a qualitative study of the public's perspective", *BMC medical ethics*, vol. 22,1 81.

les traitements de données à caractère personnel de ces personnes. Tout en affirmant le principe d'extinction au décès de la personne des droits relatifs à ses données personnelles, elle permet à la personne concernée d'anticiper la gestion de ses données personnelles. Ainsi, en vertu de l'actuel article 85 de la loi dite informatique et liberté, issu de la loi du 6 octobre 2016, « toute personne peut définir des directives relatives à la conservation, à l'effacement et à la communication de ses données à caractère personnel après son décès. Ces directives sont générales ou particulières ». La personne concernée peut désigner un tiers qui se chargera d'exécuter ses directives. Le risque est que la personne concernée ignore l'existence de ces dispositions légales. À ce propos on ne peut que saluer l'existence de documents en ligne du Groupement hospitalier universitaire de Paris[133] ou de l'Assistance Publique des Hôpitaux de Paris[134] qui informent les patients de leur droit à définir des directives sur la conservation, l'effacement et la communication de leurs données de santé après leur décès. Toutefois, on peut s'interroger sur la portée de ce droit à l'effacement car même si, comme on l'a dit, le RGPD ne s'applique pas aux personnes décédées, la loi nationale se borne à renvoyer au RGPD. Or, le droit à l'effacement prévu par l'article 17 du règlement européen n'est pas un droit absolu. Il ne peut s'exercer lorsque le traitement répond à une obligation légale ou est rendu nécessaire par l'exécution d'une mission de service public.

Par ailleurs, aux termes de l'article R. 1112-7 du code de la santé publique, les établissements de santé doivent conserver les dossiers médicaux pendant un délai de dix ans après le dernier passage du patient décédé dans l'établissement. Le code de la santé publique[135] organise la communication du dossier médical aux successeurs légaux du défunt de manière équilibrée, dans le respect de la volonté du défunt comme de la préservation du secret médical. La communication du dossier médical aux successeurs légaux n'est pas autorisée si le défunt s'y est opposé et cette communication ne s'exerce que de manière limitée : recherche des causes de la mort, exercice d'un droit ou défense de la mémoire du défunt. Les ayants droit ne seront autorisés qu'à accéder aux seuls éléments nécessaires à l'objectif poursuivi.

Il est donc souhaitable que tout entrepôt de données de santé ou plateforme explicite et informe de façon claire les personnes enregistrées de leurs droits d'effacement sur ces données après leur décès, en leur précisant l'étendue et les limites de ce droit (Recommandation n°16).

1.4 Enjeux éthiques du consentement

1.4.1 Avantages des modèles de consentement

- Le consentement rend transparentes les finalités des projets sur les PDS, ce qui renforce la confiance du public en donnant une meilleure visibilité aux projets de recherche.
- Dans le cadre des PDS, le consentement éclairé (surtout dynamique spécifique) permet aux participants à des projets de recherche d'être informés des buts poursuivis par les chercheurs et de pouvoir les comparer avec leurs propres valeurs et intérêts, et ainsi de s'engager ou non.

[133] Voir : https://www.ghu-paris.fr/fr/lentrepot-de-donnees-de-sante-eds.

[134] Voir : https://www.aphp.fr/patient-public/vos-droits/protection-des-donnees-personnelles-information-patient

[135] Voir articles L. 1117-7 et L. 1110-4 du code de la santé publique

- Le consentement dynamique, qui utilise une plateforme pour stocker les consentements, permet aux patients de se souvenir de leurs choix. Cela répond à la préconisation du CCNE dans son avis n°136 (recueil et mémoire du consentement).

L'intérêt de ce type de plateforme est illustré par l'exemple suivant. Concernant le *100,000 Genomes project*[136], qui recourt au consentement général (*broad consent*), une étude de 2020[137] montre qu'une partie des participants à l'étude n'a pas saisi les complexités du projet et les types de résultats qu'ils pouvaient induire ; par exemple, 20 % des participants au volet « cancer » interrogés ne se souvenaient pas des décisions qu'ils avaient prises concernant les découvertes secondaires.

1.4.2 Limites des modèles de consentement

- Il y a différents modèles de consentements selon les pays, ce qui pose un défi pour les projets de recherche internationaux à grande échelle.
- Le consentement en ligne pose un problème d'authentification (problème des signatures électroniques)[138].
- Il ne faut pas négliger le risque « informationnel » - c'est-à-dire la complétude de l'information – lorsqu'on veut obtenir un consentement spécifique. Comme le soulignent Mikkelsen *et al*.[139], plus que les objectifs des projets de recherche réalisés sur les PDS, ce sont les modalités de sécurisation des données sur les PDS qui peuvent inciter les personnes à consentir et à déposer leurs données. De même, un sentiment de confiance dans les porteurs de projets est fondamental pour obtenir le consentement.
- La lourdeur de la mise en place de plateformes pour le consentement dynamique doit être prise en compte : le coût peut être au détriment de la recherche, et le temps nécessaire pour obtenir le consentement considéré comme trop important par les chercheurs avant le démarrage d'un projet.
- Comme dans le cas des enquêtes par cohortes, on observe un sentiment de fatigue avec le consentement dynamique dû à la routinisation du clic (*consent fatigue*) : à solliciter trop souvent les personnes, on les lasse[140].

1.4.3 Tensions et enjeux éthiques

Le recueil des données de santé pour les PDS met à nouveau en évidence une tension entre protection des données, respect de la vie privée (consentement libre et éclairé des patients) et contribution au bien commun (progrès médical et amélioration de la santé publique).

Ce consentement est à mettre en balance avec un renoncement éventuel à la confidentialité de ses données, comme le souligne l'avis n°136 du CCNE[141].

---

[136] Voir : https://www.genomicsengland.co.uk/initiatives/100000-genomes-project

[137] Ballard L.M., Horton, R.H., Dheensa, S. *et al., (2020),* "Exploring broad consent in the context of the 100,000 Genomes Project: a mixed methods study", *Eur J Hum Genet* 28, 732–741. https://doi.org/10.1038/s41431-019-0570-7.

[138] Kogetsu, Atsushi, and Kazuto, (2022), "Framework and Practical Guidance for the Ethical Use of Electronic Methods for Communication With Participants in Medical Research", *Journal of medical Internet research* vol. 24,4 e33167.

[139] Mikkelsen R.B., Gjerris M., Waldemar G., Sandøe P., (2019), "Broad consent for biobanks is best - provided it is also deep", *BMC Med Ethics,* Oct 15;20(1):71.

[140] *Ibidem*.

[141] CCNE, avis n° 136, *op.cit.*

La demande de consentement met en évidence un enjeu d'équité. Des chercheurs[142] observent que la volonté de partager ses données et de consentir largement à des projets de recherche est inégalement répartie selon les groupes sociaux (origine ethnique, genre, niveau socio-économique), surtout avec le consentement large (*broad consent*) qui rejoint l'altruisme en matière de données dont il est question ci-après (§ III.4). Plusieurs conséquences sont à mentionner :

- Il existe des risques de biais de représentation dans les PDS avec pour conséquence un accroissement des inégalités de santé. Ce point a été abordé au § II.2.
- Des personnes susceptibles de ne pas bien comprendre les modalités et finalités des projets de recherche pourraient donner largement leur accord, même à des études douteuses, mettant ainsi en péril leurs données personnelles.
- Dans le cas du consentement dynamique, des personnes n'ayant pas les équipements pour accéder à des plateformes en ligne ou ayant une littératie numérique faible pourraient être empêchées de donner leur accord, privant éventuellement la communauté dont elles font partie des bénéfices potentiels de la recherche.

En conséquence, il faut favoriser un consentement dynamique en garantissant la confiance dans la PDS par une information régulière et transparente, sans trop de lourdeur (voir Recommandation n°17). Il faut en effet veiller à ne pas inonder d'information les patients, et à encourager la littératie numérique et l'accompagnement par des personnes de confiance qui pourraient être des auxiliaires en numérique de santé[143] (voir Recommandation n° 18).

Par ailleurs, pour que cette information puisse être transmise correctement aux patients, une formation minimale du personnel soignant, ou de la personne de confiance en tant qu'auxiliaire en numérique de santé, est nécessaire. La personne de confiance pourrait d'ailleurs recevoir cette formation minimale via les associations de patients. Il est donc crucial de mettre en place une acculturation aux enjeux techniques et éthiques des plateformes de données de santé dans les formations des soignants, tout comme dans les formations des représentants de patients et des personnes de confiance qui joueraient le rôle d'auxiliaires en numérique de santé (voir Recommandation n°19).

2. Le choix par défaut *(opt-out)*

Alors qu'un certain nombre de PDS demandent expressément un consentement pour recueillir et traiter les données personnelles des individus (stratégie ***opt-in***), d'autres ont une stratégie ***opt-out***, qui ne requiert pas l'adhésion explicite des personnes : tant qu'elles ne disent pas non, on peut considérer qu'elles sont d'accord.

2.1 En France

C'est la stratégie *opt-out* qui est appliquée concernant les données de l'Assurance-maladie. Comme rappelé sur le site Ameli[144] : « Sauf exception, vous ne pouvez pas vous opposer à ce que l'Assurance Maladie utilise vos données dans le cadre de ses missions

[142] Wiertz S., Boldt J., (2022), "Evaluating models of consent in changing health research environments", *Med Health Care Philos*, Jun;25(2): Considerations of justice, page 272.

[143] Voir : Avis commun n°141 du CCNE et n°4 du CNPEN, (Jan. 2023), *Diagnostic Médical et Intelligence Artificielle : Enjeux Ethiques*, 58 p.

[144] Voir : https://www.ameli.fr/assure/protection-donnees-personnelles

fixées par la loi ou pour des motifs de santé publique. Une information particulière est donc portée à votre connaissance via nos mentions spécifiques d'information lorsque vous disposez d'un droit d'opposition ». De son côté La CNIL précise dans une note[145] ce qu'il en est pour l'entrepôt SNDS (et donc du GIP-PDS) : « toute personne dispose d'un droit d'opposition si elle ne souhaite pas que les données qui la concernent, contenues dans le SNDS, fassent l'objet d'une utilisation à des fins de recherche. Elle ne peut toutefois pas s'opposer aux traitements de données nécessaires à l'exercice des missions des services de l'État et de certains établissements publics telles que, par exemple, le suivi d'une épidémie ou la surveillance sanitaire ».

La stratégie opt-out est aussi la stratégie adoptée par *Mon espace santé*[146], à la différence de l'initiative précédente du Dossier Médical Partagé (DMP) qu'il remplace depuis le 1er juillet 2021, et qui proposait une stratégie *opt-in*. Créé au printemps 2022, « Mon Espace Santé permet à toute personne affiliée à l'Assurance maladie de stocker et d'accéder à ses données de santé en toute confiance et en toute sécurité. (…) L'activation de cet espace santé numérique s'effectue en ligne, grâce à un code secret provisoire reçu par courriel ou par courrier. Cependant cette activation n'est en aucun cas obligatoire. Sans intervention de votre part dans les 6 semaines suivant la réception du mot de passe provisoire, votre compte sera créé automatiquement »[147]. En février 2023, 98% des assurés disposent d'un profil *Mon Espace Santé*, 7,92 millions de comptes sont activés et 26% des utilisateurs ont complété leur profil médical.[148]

2.2 Au Royaume-Uni : « National data opt-out »

Au Royaume-Uni, le *national data opt-out*[149] est un service des données nationales qui a été introduit le 25 mai 2018. Il permet aux patients de refuser que les informations confidentielles les concernant soient utilisées à des fins autres que leurs soins et traitements individuels, qu'il s'agisse de recherche (par exemple sur des médicaments pour des maladies rares) ou de planification (amélioration des services de santé et de soin). Le *national data opt-out* couvre les informations confidentielles sur les patients recueillies à propos des soins au Royaume-Uni : soins de santé et services sociaux pour adultes, financés, commandés ou coordonnés par les pouvoirs publics, et soins privés dispensés dans le cadre du *National Health Service* (NHS).

Toutes les organisations du NHS doivent fournir des informations sur le type de données qu'elles collectent et sur la manière dont elles sont utilisées. Des registres de diffusion des données sont publiés par *NHS Digital* et *Public Health England*, indiquant les données qu'ils ont partagées avec d'autres organisations.

Si les patients sont d'accord pour l'utilisation de leurs données de santé à des fins autres que leurs soins et traitements individuels, pour la recherche et la planification, ils n'ont rien à faire. Ils peuvent néanmoins consulter ou modifier leur choix d'*opt-out* des données nationales à tout moment en utilisant un service en ligne ou par un clic dans l'application du NHS. Le site en ligne du NHS[150] énumère les différents cas où, même si le patient refuse, ses données de santé pourront néanmoins être utilisées.

---

[145] CNIL, « SNDS : Système National des Données de Santé » [https://www.cnil.fr/fr/snds-systeme-national-des-donnees-de-sante].
[146] Voir : https://www.monespacesante.fr/.
[147] Voir : https://www.aide-sociale.fr/mon-espace-sante-suppression-compte/#.
[148] Voir : https://esante.gouv.fr/strategie-nationale/mon-espace-sante accédé le 13/02/2023
[149] Voir : https://digital.nhs.uk/services/national-data-opt-out.
[150] Voir : https://www.nhs.uk/your-nhs-data-matters/where-your-choice-does-not-apply/.

3. Tensions éthiques *opt-in / opt-out*

Selon le rapport de la Commission européenne sur les données de santé à la lumière du RGPD[151], certaines études[152,153] montrent que le modèle de consentement ***opt-in*** est considéré comme une pratique plus fiable de partage des données, tandis que d'autres études[154] montrent que l'*opt-out* (approche par défaut) est également acceptable si certaines conditions sont remplies.

On retrouve la tension éthique entre les deux approches, entre respect de la vie privée et service de l'intérêt général. L'approche *opt-in*, qui repose sur le consentement libre et éclairé, favorise le respect des données personnelles, éventuellement au détriment de l'alimentation des PDS, tandis que l'approche du choix par défaut favorise l'enrichissement des PDS, permettant ainsi de réduire des biais de données pour une meilleure efficacité des recherches, mais éventuellement aux dépens de l'information des patients sur l'utilisation de leurs données personnelles.

L'approche *opt-in* s'avère plus transparente. Toutefois, le consentement libre et éclairé suppose que les patients disposent d'informations suffisantes sur le plan médical et de compétences numériques minimales dans le cas d'un consentement dynamique sur une plateforme, pour donner leur consentement, ou qu'ils aient la possibilité de recourir à un tiers de confiance, capable de les aider dans cette démarche. Sans cela, il y a un enjeu d'équité.

Cette transparence peut toutefois être perçue comme lourde, voire anxiogène par les patients qui seraient régulièrement informés du lancement de nouveaux projets de recherche et sollicités pour donner ou non leur accord, selon la fréquence de cette communication.

Il convient donc de prendre en compte la tension entre le respect de la vie privée par une information nécessaire pour éclairer le consentement et le respect de la tranquillité des patients qui ne veulent pas être sur-sollicités.

4. L'altruisme en matière de données de santé

4.1 Une nouvelle forme de consentement en vue du bien commun

L'altruisme en matière de données se présente comme une seconde voie par rapport au modèle de la collecte des données régi selon le choix par défaut, et se rapproche du concept de consentement général (*broad consent*) déjà évoqué (§ III.1.2). Dans son rapport de février 2021[155], la Commission européenne mentionne plusieurs initiatives citoyennes qui ont émergé dans quelques pays, dans une démarche *bottom-up*. Il s'agit de partages de données de santé, organisés par des coopératives détenues par des personnes physiques, comme Salus-Co-op (voir annexe 4.12), ce qui leur permet de

[151] European commission, *Assessment of the EU Member States' rules on health data in the light of GDPR*, (2021), https://ec.europa.eu/health/system/files/2021-02/ms_rules_health-data_en_0.pdf

[152] Karampela M., Ouhbi S., & Isomursu M., (2019), "Connected Health User Willingness to Share Personal Health Data: Questionnaire Study", *Journal of Medical Internet Research*, 21(11).

[153] Stockdale J., Cassell J., & Ford E., (2019), ""Giving something back": A systematic review and ethical enquiry into public views on the use of patient data for research in the United Kingdom and the Republic of Ireland", *Wellcome open research*, 3, 6.

[154] Skovgaard L., Wadmann S., Hoeyer K., (2019), "A review of attitudes towards the reuse of health data among people in the European Union: The primacy of purpose and the common good", *Health Policy*, 123 (2019) 564–571.

[155] European commission, *Assessment of the EU Member States' rules on health data in the light of GDPR*, (2021), https://ec.europa.eu/health/system/files/2021-02/ms_rules_health-data_en_0.pdf, p. 113.

devenir des participants actifs dans le champ scientifique. On peut aussi considérer que les engagements volontaires de personnes dans des cohortes comme Constances et *UK Biobank* (voir § I.2.4 et annexes 4.9 et 4.10) qui mettent leurs données de santé personnelles à disposition de chercheurs, au-delà du contenu des BDS nationales standardisées, relève de l'altruisme en matière de données.

Le prêtre et expert en nouvelles technologies auprès du Vatican Eric Salobir défend l'altruisme en matière de données, avec l'idée que « protéger la donnée sans la valoriser, c'est faire seulement la moitié du chemin. La souveraineté numérique réside également dans une bonne utilisation des moyens dont nous disposons. […] Si on veut mettre les données au service de l'intérêt général, il faut être plus sélectif sur le profil des utilisateurs et plus pragmatique en trouvant les moyens de financer la gestion de ces données. C'est ce que propose le modèle du data-altruisme »[156].

Si l'altruisme en matière de données personnelles de santé est *a priori* paré de toutes les vertus pour contribuer à l'intérêt général, à la santé publique et au progrès de la recherche médicale, et s'il est d'ailleurs encouragé dans la constitution des PDS basées sur cohortes et par les associations de patients[157], on doit cependant veiller aux possibles détournements d'usage de ces données personnelles sensibles par rapport aux intentions initiales des volontaires qui les mettent à disposition. D'où l'importance du cadre juridique spécifique en cours d'élaboration.

4.2 L'altruisme dans le règlement européen sur la gouvernance des données

Afin de promouvoir la disponibilité des données et la réutilisation de certaines catégories de données protégées du secteur public, et afin de créer un environnement fiable pour faciliter leur exploitation à des fins de recherche et de création de nouveaux services et produits innovants dans l'intérêt général, la Commission Européenne a adopté le règlement sur la gouvernance des données (le *Data Governance Act* ou DGA)[158]. Ce texte introduit la notion d'altruisme en matière de données (*data-altruism*), défini comme « le partage volontaire de données, moyennant le consentement des personnes concernées au traitement de données à caractère personnel les concernant, ou les autorisations accordées par d'autres titulaires de données pour l'utilisation de leurs données à caractère non personnel sans demander ni recevoir de contrepartie »[159].

Ce partage volontaire se fait à des fins d'intérêt général, comme la santé publique, et concerne aussi bien les individus que les entreprises. Le règlement prévoit la création de deux nouveaux acteurs de la collecte de données, l'un est un prestataire de services qui s'inscrit dans le cadre commercial et l'autre est une organisation altruiste.

Le premier acteur dénommé « prestataire de service d'intermédiation de données » (articles 10 et 11) est soumis à notification à l'État membre où il a son établissement principal – contrairement à un courtier en information (*data broker*) présenté au § I.2.5. Ces prestataires de services d'intermédiation interviennent contre rémunération entre les

---

[156] Salobir E., cité dans *La Croix*, (Débat), « Faut-il partager ses données au nom de l'intérêt général ?, Partager ses données sans craindre pour sa vie privée », 28 juin 2022.

[157] A titre d'exemple on peut citer la convention multipartite signée par France Assos Santé avec le GIP-PDS, Santé publique France et Sanoïa. Voir : https://www.santepubliquefrance.fr/presse/2022/les-associations-s-engagent-pour-l-ouverture-des-donnees-de-sante-a-la-recherche-d-interet-public-une-convention-multipartite-entre-france-assos

[158] Voir : https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R0868. Publié au Journal Officiel de l'Union Européenne le 23 juin 2022, le DGA entrera en vigueur en septembre 2023.

[159] Voir : https://www.europarl.europa.eu/doceo/document/A-9-2021-0248_FR.html.

détenteurs de données et les utilisateurs potentiels de ces données, y compris par la mise à disposition des moyens, techniques ou autres, nécessaires pour permettre la fourniture de ces services. Ce nouveau régime de prestataires de services d'intermédiation définit les conditions liées à la fourniture de ces services et leur interdit d'utiliser les données pour une autre finalité que la mise à disposition des utilisateurs, ce qui rappelle le principe de limitation des finalités des traitements des données personnelles. Il s'agit d'instaurer la confiance et de garantir la neutralité de ces organismes. A titre d'exemple de ce type de structures, en dehors du domaine de la santé, on peut citer la plateformes d'échange de données Dawex (voir Annexe 4.11) en France ou bien le *Data Intelligence Hub* de Deutsche Telekom[160] en Allemagne.

La création de ce second nouvel acteur de données s'inscrit dans le cadre de l'élaboration par les États de politiques nationales dans le domaine de l'altruisme en matière de données (article 16). Si dans ce cadre, des personnes acceptent de mettre à disposition volontairement, par altruisme, des données à caractère personnel les concernant détenues par des organismes du secteur public, il faut que les organisations altruistes en matière de données qui vont se créer sur ce fondement pour collecter ces données pour des objectifs d'intérêt général soient indépendantes et à but non lucratif**.** Ces organisations altruistes en matière de données devront, pour être reconnues comme telles, être inscrites dans un registre qui sera tenu au niveau national et européen, ce qui leur permettra d'être reconnues dans toute l'Union européenne. L'article 21 prévoit des garanties très strictes pour les personnes à l'origine des données personnelles. Les organisations altruistes ne doivent pas utiliser les données pour des objectifs autres que ceux d'intérêt général pour lesquels la personne concernée ou le détenteur des données autorise le traitement. Elles ne doivent pas recourir à des pratiques commerciales trompeuses pour solliciter la fourniture de données. Elles fournissent des outils permettant d'obtenir le consentement des personnes concernées ou l'autorisation de traiter des données mises à disposition par des détenteurs de données ainsi que des outils permettant de retirer facilement ce consentement ou cette autorisation.

Le règlement prévoit aussi la création d'un formulaire de recueil de consentement (article 25), qui sera valable dans toute l'Union européenne comme modèle de consentement pour le partage des données et leur réutilisation, afin d'accroître la transparence à l'égard des personnes concernées et d'instaurer la confiance nécessaire pour encourager les particuliers et les entreprises à transmettre leurs données à ces organisations.

4.3 Vigilance sur l'altruisme en matière de données

La CNIL et ses homologues européennes ont souligné la nécessité de veiller à la cohérence des deux textes européens, *Digital Governance Act* et *Digital Act,* avec le RGPD et réclamé « une gouvernance intelligente gravitant autour des autorités de protection des données afin d'assurer l'application efficace et effective des différents cadres juridiques et d'assurer leur lisibilité pour les personnes et acteurs économiques concernés »[161]. Toutefois, même s'il est bien spécifié que les modalités de consentement à l'altruisme en matière de données doivent être conformes au RGPD et qu'en cas de conflit avec le règlement sur la gouvernance des données, le RGPD prévaut, il n'est pas aisé de faire respecter ses droits, ni même de savoir s'ils sont respectés[162]. Une fois qu'une personne

[160] Voir : https://dih.telekom.com

[161] CNIL, « Stratégie européenne pour la donnée : la CNIL et ses homologues se prononcent sur le Data Governance Act et le Data Act » [https://www.cnil.fr/fr/strategie-europeenne-pour-la-donnee-la-cnil-et-ses-homologues-se-prononcent-sur-le-data-governance].

[162] Gonzales Fuster G., citée dans *La Croix*, (Débat), *Faut-il partager ses données au nom de l'intérêt général ? Des risques de pratique discriminatoires,* 28 juin 2022.

a donné son consentement, il semble qu'elle n'ait pas de visibilité sur ce que deviennent ses données, ce qui nous semble regrettable (voir Recommandation n°17).

Par ailleurs, l'altruisme en matière de données peut conduire à une mauvaise représentativité des données, puisque le contenu de la plateforme les recueillant dépend du bon vouloir des personnes qui l'alimentent. Bien que ce risque de biais des données pour les PDS (voir § I.2.1) ne soit pas spécifique à l'altruisme en matière de données, il est certainement accentué dans ce cas. Pour redresser ces biais de représentativité de la population on peut recourir à la constitution d'un « échantillon témoin », comme le propose la cohorte Constances[163], en demandant à des non-participants à Constances de fournir leurs données de l'Assurance Maladie et de la Caisse nationale d'assurance vieillesse, lesquelles seront traitées anonymement, afin de revenir à une meilleure représentativité.

Mentionnons enfin deux risques de l'altruisme en matière de données soulignés par le rapport ITF-Sopra-Steria Next[164] : (i) le *green-washing* (ou *bluewashing*) : une entreprise pourrait mettre en avant une démarche altruiste en matière de données pour masquer des comportements dommageables qu'elle a pu avoir et qui ont été pointés par les médias et (ii) la tromperie : le Bureau européen des unions de consommateurs rappelait son inquiétude « sur la façon dont une définition floue de l'altruisme [...] pourrait permettre aux entreprises d'abuser de motifs vagues et altruistes pour pousser les consommateurs à partager leurs données »[165], diverses formes de hameçonnage étant envisageables. D'où la nécessité de transparence.

Le recours à l'altruisme en matière de données repose sur trois facteurs : (i) la confiance des donateurs en ceux qui auront accès à leurs données, (ii) la sensibilité plus ou moins grande des données, sensibilité très grande concernant les PDS et (iii) dans quelle mesure ces données peuvent contribuer au bien public. Parmi les freins et réticences figure le fait que les données soient utilisées à des fins commerciales ou par des organisations gouvernementales.

Enfin, ce modèle d'altruisme des données ne prend pas suffisamment en compte la valeur réelle d'accès et d'utilisation des données des PDS, et le coût pour celles-ci, comme, par exemple, pour la PDS de l'AP-HP. De plus, aucun bénéfice n'est prévu en retour pour les personnes qui auront donné par altruisme accès à leurs données, par exemple en matière de résultats des recherches. Sur ce point, le rapport HTF-Sopra Steria[166] précise que « un modèle data-altruiste peut inclure (i) des mécanismes de financement afin d'assurer la mise en place et la maintenance des infrastructures de partage (ii) des incitations à participer aux modèles, qu'elles soient directes ou indirectes ». Il ajoute que : « Si le DGA prévoit d'ores et déjà que les tiers de confiance peuvent instaurer des systèmes d'accès aux données qu'ils détiennent en contrepartie de redevances, il convient de penser à l'instauration de systèmes d'incitations directes ou indirectes au bénéfice des contributeurs de données »[167]. Sur ce dernier point, nous préconisons la reconnaissance des contributeurs de données de santé (voir Recommandation n° 14).

[163] Voir : https://www.constances.fr/espace-volontaires/representativite.php

[164] Rapport HTF – Sopra-Steria Next, (2022), « Data-altruisme, une initiative européenne. Les données au service de l'intérêt général », Rapport Human Technology Foundation, et Exploratoire Sopra-Steria Next, p. 28.

[165] L. Bertuzzi, "Data Governance: new EU law for data-sharing adopted", *Euratic*, 1er décembre 2021. Cité dans le rapport HTF-Sopra-Steria, *op. cit.* p 41.

[166] Rapport HTF-Sopra Steria, *op. cit.*

[167] *Ibidem*.

5. Pour un écosystème collaboratif pour les plateformes de données de santé

Nous rassemblons un certain nombre de conditions que la conception et la gestion des plateformes de données de santé doivent satisfaire pour assurer un écosystème de santé respectant les principes éthiques que nous avons développés dans le présent avis. Elles vont dans le sens de la synthèse sur la consultation citoyenne TEHDAS[168], qui conclut que « la bienfaisance conditionnelle des citoyens à l'égard de la réutilisation des données de santé nécessite un cadre bien pensé, accordant une attention suffisante à ses dimensions éthiques, juridiques et sociétales ».

5.1 Garantir le bien commun

La tension éthique entre la protection des données de santé, le respect de la vie privée (consentement libre et éclairé des patients) et la contribution au bien commun par le partage des données (progrès médical et amélioration de la santé publique) soulignée au § III.1.4 pose la question des contours du bien commun en matière d'utilisation des données de santé. Une difficulté résulte du fait que ce bien commun peut varier selon l'information, les intérêts et préférences des personnes. Cela devrait amener à impliquer les citoyens dans la gouvernance des PDS de façon efficiente. Cela leur permettrait de s'exprimer sur les modalités du consentement à mettre en place et sur leurs besoins en recherche (voir Recommandation n°21).

Cette implication ne va pas de soi.[169] En effet, plusieurs enquêtes dont celle menée par la fondation Roche en 2021 montrent les fortes inégalités d'accès au numérique par les individus, et partant, de leur conscience des risques et des bénéfices associés à ces données[170]. Le citoyen isolé aura ainsi tendance à négliger l'importance de son implication dans les PDS, aussi bien au niveau de la conception de l'infrastructure que de sa gouvernance.

5.2 Favoriser la participation citoyenne à la gouvernance des plateformes de données de santé via les associations

En revanche, lorsque les patients sont organisés, par exemple dans une association de patients ou par l'intermédiaire d'un institut, ils perçoivent alors beaucoup mieux l'importance du rôle qu'ils sont susceptibles de jouer et s'investissent donc davantage. L'exemple de l'association de patients atteints de maladies rénales Renaloo est à ce titre très éclairant. Yvanie Caillé, membre fondatrice, a montré l'intérêt à participer à la gouvernance en devenant directrice l'Institut National aux Données de Santé. De même, une instance de la gouvernance de la plateforme des données en cancérologie de l'Institut national du cancer (INCa) (annexe, Annexe 4.7) est son comité scientifique et éthique où siègent des représentants de patients. La participation citoyenne à la gouvernance des PDS est bien plus efficace lorsqu'elle passe par l'intermédiaire d'organisations de patients.

Plus généralement, les enjeux éthiques relatifs à l'écosystème de santé au cœur des PDS amènent à développer des processus de concertation et d'information citoyennes à toutes les étapes de la chaîne de traitement des données de santé : en amont lors de la définition des projets de recherche et des décisions de valorisation des recherches, pendant leur

[168] TEHDAS : Towards European Health Data Space, (2022), *Le Débat des Données, une consultation citoyenne en ligne sur la réutilisation des données de santé – Rapport intermédiaire*, 36 p.

[169] Fréderic Graber, *L'inutilité publique ne va pas de soi, Histoire d'une culture politique française*, Paris, éditions Amsterdam, 2022, 208 p.

[170] Voir : https://fondationroche.org/wp-content/uploads/sites/8/2021/10/fondation-roche_rapport-observatoire-acces-numerique-2021.pdf

réalisation, et après, pour être tenus informés des résultats obtenus (voir Recommandation n°20).

5.3 Construire la confiance par l'information, la transparence, la formation et l'accompagnement numérique

Concernant l'information des patients, nous préconisons la rédaction d'un livret ou d'une fiche d'information disponible dans les salles d'attente des hôpitaux ou des cabinets médicaux en ville, tout comme dans les officines de pharmacie, informant les personnes de l'usage qui est fait par les praticiens de leurs données (voir Recommandation n°18). C'est fondamental, en particulier, pour les informations relatives à *Mon espace santé* qui se déploie depuis début 2022, attendu que les soignants ne signalent pas aux patients s'ils transfèrent sur cet espace les données qu'ils recueillent sur eux, pas plus qu'ils ne leur demandent l'autorisation d'accéder à *Mon espace santé*, lorsque celui-ci est activé par les patients.

Rappelons qu'à ce titre, les soignants doivent avoir la capacité de formuler oralement cette information, ce qui nécessite une formation minimale aux enjeux techniques et éthiques, comme préconisé plus haut (voir § III.1.4 et Recommandation n°19).

La circulation d'une information claire et loyale contribue à la bonne qualité des relations interpersonnelles entre soignants et soignés, et favorise le partage des données entre patients, soignants et administrateurs de plateformes de stockage.

Si les patients attendent beaucoup de transparence de la part des soignants, il faut souligner le fait que les données recueillies par les soignants reflètent leurs pratiques médicales et les exposent aux regards des patients qui y ont accès. Il s'agit bien d'établir une relation de confiance réciproque entre soignants et soignés.

Les conditions attendues d'un tel partage en confiance d'un point de vue éthique sont loyauté, bienveillance, non surveillance, autonomie des patients.

Cette confiance repose sur des propriétés techniques pour les PDS, comme la sécurité, l'interopérabilité, la portabilité, l'anonymisation ou la pseudonymisation des données pour des usages bien définis (cf § I.4), la souveraineté (cf § II.1), et sur la garantie d'une utilisation conforme au RGPD, qui préserve la confidentialité des données et leur usage pour le bien commun. Elle se consolide par un écosystème de concertation, où prévaut la transparence des usages des PDS, montrant les avantages de leur conception et utilisation, sans cacher leurs limites et les éventuels risques qu'elles comportent. Cette même transparence est attendue pour les projets de recherche, où les finalités poursuivies doivent être clairement annoncées, et les grands domaines de recherche devraient pouvoir être décidés en concertation. Le devenir des données pour un usage primaire alimentant les PDS et des données issues des travaux de recherche doit être explicité, et leur valorisation doit faire l'objet d'une concertation. Enfin, il faut garder à l'esprit que le développement de ces PDS soulève le problème de l'égalité d'accès aux données de santé des personnes, en raison des inégalités en termes de littératie numérique.

6. Recommandations

- **Recommandation n°16 :** Informer clairement les patients qu'ils ont le droit de définir des directives relatives à la conservation, à l'effacement et à la communication des données personnelles les concernant contenues dans une PDS après leur décès.

- Recommandation n°17 : Promouvoir une forme consentement qui préserve le lien entre la personne qui donne son consentement et celle qui le reçoit, de façon que la personne puisse consentir en confiance à des types de projets (et non pas des projets spécifiques) en étant informée de manière transparente et régulière sur les projets et les partenariats qui utiliseront ses données.
- Recommandation n°18 : Développer une information des personnes sur l'usage de leurs données de santé et sur les PDS existantes, qui soit adaptée à leur culture numérique par divers canaux et dans les lieux qu'elles fréquentent (hôpitaux, pharmacies, cliniques privées) et proposer un accompagnement par des tiers de confiance qui pourraient être des auxiliaires en numérique de santé.
- Recommandation n°19 : Créer des formations répondant aux besoins de nouvelles compétences médicales et numériques concernant les PDS pour les personnels soignants, les informaticiens, les utilisateurs de PDS, et les auxiliaires en numérique de santé.
- Recommandation n° 20 : Accompagner la mise en œuvre de l'altruisme en matière de données de santé par une information régulière et transparente des personnes qui mettent leurs données à disposition sur les usages qui sont faits de leurs données.
- Recommandation n°21 : Favoriser la participation citoyenne à la gouvernance des PDS, et à l'élaboration des appels à projets de recherche, en particulier par l'intermédiaire des associations de patients, et informer ces dernières en amont, pendant et après les projets de recherche.

ANNEXES

Annexe 1 : Recommandations

Dans les recommandations qui suivent, nous abrégeons « Plateforme de données de santé » par PDS.

Sur la qualité et le partage des données

- **Recommandation n°1** : Expliciter la nature et l'origine des données personnelles de santé rassemblées dans les PDS, en distinguant leurs usages primaires et leurs usages secondaires et, pour un projet de recherche donné, utiliser des ensembles de données non biaisées, ou, lorsque cela n'est pas possible, tenir compte de ces biais dans leur analyse, par exemple, par des méthodes de pondération.
- **Recommandation n°2 :** Veiller à ce que la durée de conservation des données de santé publique collectées soit bien calibrée par rapport aux exigences de la recherche sans négliger la nécessaire protection des données personnelles.

Sur l'impact environnemental des PDS

- **Recommandation n°3** : Evaluer l'impact environnemental des PDS et viser leur sobriété énergétique par des choix appropriés de stockage des données, d'architecture, et de modes de fonctionnement.

Sur l'architecture des PDS :

- **Recommandation n°4** : Demander aux pouvoirs publics de s'impliquer davantage dans l'élaboration de standards et de normes pour formater et structurer les données de santé afin de favoriser de meilleures portabilité et interopérabilité des PDS.
- **Recommandation n°5** : Mener des études d'évaluation comparatives entre les approches centralisées et les approches décentralisées des PDS, et sur leurs combinaisons, pour assurer une gestion sécurisée des données de santé. Encourager les innovations en Intelligence Artificielle fédérée pour informer le débat entre architectures centralisées et décentralisées.
- **Recommandation n°6 :** Choisir des solutions d'architecture de PDS qui respectent les écosystèmes locaux et tiennent compte de projets de recherche multicentriques qui nécessitent des données réparties dans divers centres cliniques ou hôpitaux, mettant en évidence l'intérêt de mutualiser les données.
- **Recommandation n°7** : Inciter les créateurs de PDS publiques à adopter des formats standards ouverts et des algorithmes *open source* pour mettre en qualité les données et traiter ultérieurement des flux de données, et aussi permettre des études multicentriques, afin de libérer le potentiel d'innovation de tous les réutilisateurs de données de santé.

Sur l'anonymisation :

- **Recommandation n°8** : Développer la recherche sur les méthodes alternatives à l'anonymisation et à la pseudonymisation des données, notamment les techniques de chiffrement homomorphe, pour pouvoir mieux exploiter des données de santé.

Sur la souveraineté :

- Recommandation n°9 : Exiger pour la création et la certification d'une PDS une démarche pluridisciplinaire, avec des experts en santé, en informatique, en droit et en sciences humaines et sociales pour anticiper les risques d'atteinte à la souveraineté.
- Recommandation n°10 : Dans les contrats de partenariat internationaux impliquant des données de santé, veiller à ce qu'il y ait des clauses garantissant que les acteurs non européens respectent les principes fondamentaux du RGPD, du règlement sur la gouvernance des données, et du futur règlement européen sur les données, pour protéger les données sensibles.
- Recommandation n°11 : Prévoir systématiquement des clauses spécifiques de transparence, d'explicabilité et de réversibilité, pour les entreprises, notamment extra-européennes, permettant en particulier des transferts de données à des coûts minimaux.
- Recommandation n°12 : Veiller aux conditions d'accès aux données de santé et d'exportation temporaire lors de la maintenance des systèmes informatiques servant les PDS impliquant des acteurs non européens.

Sur la valorisation des données :

- Recommandation n°13 : Encourager un financement des PDS sur la base de leurs coûts d'investissement et de fonctionnement, et une tarification adaptée à différents usagers, en particulier pour la recherche scientifique d'intérêt public.
- Recommandation n°14 : Inciter les entreprises qui rencontrent le succès financier en partie grâce à des jeux de données fournis par des plateformes publiques de données de santé, à partager une partie de leurs bénéfices avec ces dernières, par la signature volontaire d'une charte engageant leur réputation.
- Recommandation n°15 : Veiller à maintenir une indépendance forte entre les équipes de direction des plateformes publiques de données de santé et celles des entreprises utilisatrices afin de prévenir les conflits d'intérêts.

Conditions pour un écosystème collaboratif pour les PDS :

- Recommandation n°16 : Informer clairement les patients qu'ils ont le droit de définir des directives relatives à la conservation, à l'effacement et à la communication des données personnelles les concernant contenues dans une PDS après leur décès.
- Recommandation n°17 : Promouvoir une forme consentement qui préserve le lien entre la personne qui donne son consentement et celle qui le reçoit, de façon que la personne puisse consentir en confiance à des types de projets (et non pas des projets spécifiques) en étant informée de manière transparente et régulière sur les projets et les partenariats qui utiliseront ses données.
- Recommandation n°18 : Développer une information des personnes sur l'usage de leurs données de santé et sur les PDS existantes, qui soit adaptée à leur culture numérique par divers canaux et dans les lieux qu'elles fréquentent (hôpitaux, pharmacies, cliniques privées) et proposer un accompagnement par des tiers de confiance qui pourraient être des auxiliaires en numérique de santé.
- Recommandation n°19 : Créer des formations répondant aux besoins de nouvelles compétences médicales et numériques concernant les PDS pour les personnels

soignants, les informaticiens, les utilisateurs de PDS, et les auxiliaires en numérique de santé.

- Recommandation n° 20 : Accompagner la mise en œuvre de l'altruisme en matière de données de santé par une information régulière et transparente des personnes qui mettent leurs données à disposition sur les usages qui sont faits de leurs données.
- Recommandation n°21 : Favoriser la participation citoyenne à la gouvernance des PDS, et à l'élaboration des appels à projets de recherche, en particulier par l'intermédiaire des associations de patients, et informer ces dernières en amont, pendant et après les projets de recherche.

Pour la recherche et l'innovation :

Parmi ces recommandations, certaines concernent plus particulièrement la recherche et l'innovation. Il s'agit de la Recommandation n°2 sur la durée de stockage des données pour la recherche, de la Recommandation n°5 sur la recherche en matière l'Intelligence Artificielle fédérée et de la Recommandation n°8 sur les méthodes alternatives à l'anonymisation et à la pseudonymisation des données.

Annexe 2 : Membres du groupe de travail

Groupe de travail conjoint issu du Comité Consultatif National d'Éthique (CCNE) pour les sciences de la vie et de la santé et du Comité National Pilote d'Éthique du Numérique (CNPEN).

Gilles Adda (CCNE&CNPEN)

Thomas Bourgeron (CCNE)

Jacques Bringer (Invité extérieur - ERE Occitanie)

Sophie Crozier (CCNE)

Pierre Delmas-Goyon (CCNE)

Emmanuel Didier (CCNE) Rapporteur

Christine Froidevaux (CNPEN) Rapporteure

Fabrice Gzil (CCNE)

Jeany Jean-Baptiste (CNPEN)

Claude Kirchner (CCNE&CNPEN)

Caroline Martin (CCNE&CNPEN)

Jérôme Perrin (CNPEN) Rapporteur

Valéry Ravix (Invité extérieur - ERE PACA-Corse)

Camille Darche (rédactrice)

Hanna le Derrien (stagiaire)

Lucie Guimier (rédactrice)

Anaëlle Martin (rédactrice)

Amélie Turci (stagiaire)

Annexe 3 : Risques juridiques sur le transfert des données aux États-Unis

Le juge des référés du Conseil d'État a, le 13 octobre 2020, rejeté une requête tendant à la suspension de la Plateforme de données de santé (GIP-PDS) dite *Health Data Hub*[171]. La requête faisait état des risques de transfert de données vers les États-Unis.

Le juge des référés s'est d'abord prononcé au regard des stipulations contractuelles et a relevé que la Plateforme des données de santé (GIP-PDS) et la filiale irlandaise de Microsoft se sont engagées, par contrat d'avril 2020, à refuser tout transfert de données de santé, alors stockées aux Pays-Bas, en dehors de l'Union européenne, que Microsoft ne traitera pas les données de la Plateforme en dehors de la zone géographique spécifiée par celle-ci sans son approbation et que dans l'hypothèse où un accès aux données serait nécessaire pour les besoins des opérations d'exploitation des services en ligne et de résolution d'incidents menées par Microsoft depuis un lieu extérieur à cette zone, il serait soumis à l'autorisation préalable de la Plateforme qui s'est engagée à ne pas l'accorder. Il a ajouté qu'un arrêté ministériel du 9 octobre 2020 interdit tout transfert de données à caractère personnel en dehors de l'Union européenne dans le cadre de ce contrat. Le juge des référés a d'ailleurs demandé à la Plateforme de préciser dans un nouvel avenant qu'elle s'interdisait d'autoriser tout transfert de traitement, ce qui permet d'introduire dans le contrat l'arrêté du 9 octobre 2020. On peut en déduire que tout transfert de données même pour une opération de maintenance est interdit.

Mais ces garanties contractuelles ne lui sont pas apparues suffisantes face à la loi américaine. Le juge des référés a relevé qu'il ne peut être totalement exclu que les autorités américaines, dans le cadre de programmes de surveillance et de renseignement, demandent à Microsoft et à sa filiale irlandaise l'accès à certaines données, et que si ce risque ne justifiait pas, à très court terme, la suspension de la Plateforme, il imposait de prendre des précautions particulières, sous le contrôle de la CNIL.

Une demande de transfert pourrait en effet être faite par les États-Unis sur deux fondements juridiques : l'article 702 du FISA et le décret présidentiel EO 12333, alors même que ces données sont hébergées sur le territoire de l'Union européenne et que les termes du contrat conclu entre la Plateforme des données de santé et Microsoft s'y opposeraient.

En premier lieu, le juge des référés a relevé que les mesures techniques mises en œuvre par Microsoft ou susceptibles de l'être à brève échéance n'écartent pas toute possibilité pour cette entreprise d'accéder aux données traitées sous la responsabilité de la Plateforme des données de santé, en dépit des précautions, limitant ce risque, qui entourent le chiffrement dont elles font l'objet et le stockage des clés de chiffrement utilisées. Microsoft pourrait donc accéder aux données de la plateforme.

En second lieu, face à l'existence d'un risque, et compte tenu du fait que le juge des référés ne peut prononcer que des mesures de très court terme, il a demandé à la Plateforme de continuer, sous le contrôle de la CNIL, à travailler avec Microsoft pour renforcer la protection des droits des personnes concernées sur leurs données personnelles dans l'attente d'une solution qui permettra d'éliminer tout risque d'accès aux données personnelles par les autorités américaines, comme annoncé par le secrétaire d'État au numérique le jour même de l'audience au Conseil d'État (choix potentiel d'un nouveau sous-traitant, recours à un accord de licence suggéré par la CNIL…). Le juge des référés a estimé nécessaire de poursuivre l'exploitation pour les besoins de la gestion de l'urgence sanitaire et de l'amélioration des connaissances sur le SARS-CoV-2.

[171] Voir : https://www.legifrance.gouv.fr/juri/id/CETATEXT000042444915.

Au-delà de cette décision rendue dans l'urgence et qui ne préempte pas juridiquement l'avenir, on se reportera à l'arrêt de la CJUE dit Schrems II (C-311/18) du 16 juillet 2020. La Cour, qui ne s'est pas interrogée sur la question spécifique que posent les données collectées par le GIP-PDS, car la question n'était pas en litige, a examiné la question de manière générale.
S'agissant de l'article 702 du FISA, elle juge qu'il ne fait ressortir d'aucune manière l'existence de limitations à l'habilitation qu'il comporte pour la mise en œuvre des programmes de surveillance aux fins du renseignement extérieur, pas plus que l'existence de garanties pour des personnes non-américaines potentiellement visées par ces programmes.

S'agissant des programmes de surveillance fondés sur l'E.O. 12333, la Cour a jugé que, selon les éléments dont elle disposait, ce décret ne confère pas non plus de droits opposables aux autorités américaines devant les tribunaux.

S'agissant des deux programmes, elle a jugé qu'il n'est pas permis de considérer que les programmes de surveillance fondés sur ces dispositions sont limités au strict nécessaire.

S'agissant de l'obligation de protection juridictionnelle, elle a aussi jugé qu'aucun des deux textes n'assurait un recours effectif devant un tribunal indépendant et impartial dans le respect des conditions prévues à l'article 47 de la Charte des droits fondamentaux.

Précisons que la société Microsoft a soutenu que le GIP-PDS ne relève pas de cet arrêt car l'article 702 du FISA comme l'E.O 1233 ne pourraient servir, selon la société, de fondement légal à la surveillance de la plateforme de santé pour divers motifs.[172] Il s'agit d'une question juridique réelle sur laquelle il n'existe pas de jurisprudence. Même si Microsoft avait raison, l'opinion publique et les défenseurs des libertés individuelles conserveraient toujours un soupçon et une inquiétude liés aux pratiques américaines révélées par E. Snowden. Il est préférable de veiller à garantir une réelle autonomie dans le domaine des données de santé, avant d'attendre une solution jurisprudentielle, d'autant plus que les lois et réglementations américaines peuvent toujours être durcies.

[172] Assemblée nationale Mission d'information de la Conférence des Présidents, « Bâtir et promouvoir une souveraineté numérique nationale et européenne » audience du jeudi 27 mai 2021.

Annexe 4 : Exemples de structures proposant des services de données de santé

1. SNDS

Géré par la Caisse Nationale de l'Assurance Maladie des Travailleurs Salariés (CNAMTS), et créé en 2016, le SNDS[173,174] est un entrepôt de données médico-administratives pseudonymisées couvrant l'ensemble de la population française et contenant l'ensemble des soins présentés au remboursement. Il permet de chaîner des données provenant de différentes bases de données. Il rassemble[175] :

- les données de l'Assurance Maladie (base du système national d'information interrégimes de l'Assurance Maladie – base de données Sniiram) ;
- les données des hôpitaux (base du programme de médicalisation des systèmes d'information – PMSI) historiquement appariées au Sniiram ;
- les bases de données sur les causes médicales de décès (base du centre d'épidémiologie sur les causes médicales de décès de l'Institut national de santé et de la recherche médicale – CépiDc de l'Inserm) ;
- les données relatives au handicap (en provenance des maisons départementales des personnes handicapées – MDPH - données de la caisse nationale de solidarité pour l'autonomie - CNSA).

L'accès aux données du SNDS est très réglementé[176] et leur utilisation ne peut se faire que dans des conditions respectant le référentiel de sécurité, visant à garantir la confidentialité et l'intégrité des données et la traçabilité des accès et autres traitements. Pour cela, chaque patient est repéré dans les bases du SNDS par un pseudonyme, obtenu par l'application au NIR (numéro d'inscription au répertoire national d'identification des personnes physiques) d'un procédé cryptographique irréversible appelé FOIN (Fonction d'occultation des identifiants nominatifs).

À noter que le SNDS a pour but de rassembler et de fournir des données mais ne propose pas d'infrastructure logicielle ou de capacités de calculs pour effectuer des traitements. Il existe à côté du SNDS des structures privées, comme HEVA[177], dont le but est d'exploiter les données du SNDS dans une bulle sécurisée, sans les stocker.

Par un décret de juin 2021 le GIP-PDS et la CNAM ont été désignés comme responsables conjoints de traitement du SNDS [178]. Par ce même décret Le consortium AgorIA Santé a été lancé en juin 2021 par Docaposte, AstraZeneca et Impact Healthcare. Il bénéficie d'une autorisation de la CNIL délivrée le 23 mai 2022 pour constituer son entrepôt de données de santé avec un système fils du SNDS, une première pour un consortium d'acteurs privés.

[173]Voir : https://www.snds.gouv.fr/SNDS/Qu-est-ce-que-le-SNDS.

[174]Voir : https://documentation-snds.health-data-hub.fr/introduction/01-snds.html#les-donnees-presentes-et-absentes.

[175] Voir : https://assurance-maladie.ameli.fr/etudes-et-donnees/presentation-systeme-national-donnees-sante-snds.

[176] L'usage des données du SNDS est interdit pour i) La promotion des produits de santé, en direction des professionnels de santé ou des établissements de santé, et ii) l'exclusion de garanties des contrats d'assurance ou la modification de cotisations ou de primes d'assurance pour un individu ou un groupe d'individus. https://www.snds.gouv.fr/SNDS/Finalites-autorisees

[177] Voir : https://hevaweb.com/fr/.

[178] Voir : https://www.dsih.fr/article/4763/la-creation-d-un-entrepot-de-donnees-de-sante-par-un-consortium-d-acteurs-prives-autorisee-par-la-cnil.html.

2. GIP-PDS

Le Groupement d'intérêt public « Plateforme des données de santé » (GIP-PDS) communément appelé *Health Data Hub* a été créé par la loi du 24 juillet 2019 relative à l'organisation et la transformation du système de santé. Sa structure de groupement d'intérêt public (GIP) associe 56 parties prenantes, en grande majorité issues de la puissance publique (CNAM, CNRS, Haute Autorité de santé, France Assos Santé, directions des ministères etc.). Le GIP-PDS met en œuvre les grandes orientations stratégiques relatives au Système National des Données de Santé (SNDS) fixées par l'État et notamment le ministère des Solidarités et de la Santé. Son financement est majoritairement public. Articulées autour de quatre enjeux stratégiques : i) mettre en valeur le patrimoine de données de santé, ii) faciliter l'usage des données de santé, iii) protéger les données de santé des personnes, iv) innover avec l'ensemble des acteurs de l'écosystème, les offres de service du GIP-PDS visent à créer un véritable renfort de capacités à innover pour faire de la France un leader de l'analyse des données de santé.

3. Ouest Data Hub

En 2020 le Groupement de Coopération Sanitaire (GCS) HUGO qui rassemblait cinq Centres Hospitaliers Universitaires (CHU) du Grand Ouest : Angers, Brest, Nantes, Rennes, Tours et l'Institut de cancérologie de l'Ouest a franchi une étape majeure avec le lancement du « Ouest DataHub », la première plateforme de données hospitalières en Europe.

Celle-ci permet un regroupement des données anonymisées des six établissements membres au service de la recherche médicale, un moyen innovant d'imaginer de nouveaux projets de recherche, de développer la médecine personnalisée par des outils d'aide à la décision pour les cliniciens et les patients, et d'améliorer la vigilance sanitaire à l'échelle du Grand Ouest.

4. Entrepôt de l'AP-HP

L'EDS de l'AP-HP[179] est présenté comme le premier entrepôt de données hospitalier d'Europe. Il a pour but de rassembler, standardiser et structurer les données administratives et cliniques, les comptes-rendus d'hospitalisation, les prescriptions, les résultats d'examens biologiques et d'imagerie de plus de 13 millions de patients pris en charge par le CHU (39 établissements), avec des données remontant pour certaines à 2012. Un tableau mis à jour en 2021[180] indique les utilisations pour la recherche des différents ensembles de données constituant l'entrepôt. L'EDS a été autorisé par la CNIL en 2017. L'EDS de l'AP-HP est hébergé dans un *cloud* privé sécurisé[181] et s'appuie sur des capacités de calcul adaptées aux jeux de données et aux algorithmes mobilisés avec 20 cartes GPU *(graphical processing units)*. Pour le pilotage, l'accès aux données et aux solutions d'informatique décisionnelle (*Business Intelligence*) est assuré sur le portail PILOTE, adossé à la technologie IBM COGNOS, qui ouvre des possibilités de suivi et d'analyse de l'activité par les équipes de soins, les directions stratégiques et les groupes hospitalo-universitaires. L'EDS de l'AP-HP a mis en place une plateforme Big Data[182] qui permet l'exploitation de ses données. Le portail JUPYTER offre des espaces privés et sécurisés de traitement des données à l'aide des langages informatiques usuels[183]. En ce

[179] Voir : https://eds.aphp.fr/nos-services/eds-donnees.
[180] Voir : https://eds.aphp.fr/sites/default/files/2021-09/EDS_Disponibilite_des_donnees_20210910.pdf.
[181] Structures propriétaires certifiées hébergeurs données de santé, dans les locaux de l'AP-HP.
[182] Voir : https://eds.aphp.fr/nos-services/plateforme-outils.
[183] Voir : https://www.aphp.fr/connaitre-lap-hp/recherche-innovation/lentrepot-de-donnees-de-sante-de-lap-hp

sens, l'EDS de l'AP-HP est plus qu'un entrepôt de données et a les caractéristiques d'une PDS. Suite aux recommandations du CCNE pour sa gouvernance, l'EDS de l'AP-HP s'est doté d'un Comité scientifique et éthique (*Institutional review board*) qui comprend des représentants de patients.

5. CASD

Actuellement plutôt méconnu des scientifiques, notamment du secteur public, le CASD (Centre d'Accès Sécurisé aux Données)[184] créé par arrêté interministériel du 29 décembre 2018, est un groupement d'intérêt public rassemblant l'État représenté par INSEE, le GENES, le CNRS, l'École polytechnique et HEC. Il a pour objet principal « d'organiser et de mettre en œuvre des services d'accès sécurisé pour les données confidentielles à des fins non lucratives de recherche, d'étude, d'évaluation ou d'innovation, activités qualifiées de « services à la recherche », principalement publiques. Il a également pour mission de valoriser la technologie développée pour sécuriser l'accès aux données dans le secteur privé. En pratique, le CASD est un tiers de confiance entre le producteur et l'utilisateur de données personnelles qui assure un stockage et un accès tout à la fois sécurisé et conforme à la réglementation européenne (RGPD).
Le CASD permet donc d'avoir accès à des données sensibles au sens du RGPD, en particulier à des fins de recherche. C'est une solution nécessitant de mettre en œuvre des protocoles d'accès spécifiques et contrôlés aux données et à leur manipulation.

6. Mon espace santé

Mon espace santé[185] est d'abord une base de données qui permet de trier des informations et d'utiliser une messagerie sécurisée ; mais comme elle offre également un catalogue de services (disponibles depuis novembre 2022) on peut considérer que c'est une PDS. Le responsable de traitement et le sous-traitant sont la CNAM et le Ministère chargé de la Santé. Les données sont hébergées en France par deux sous-traitants : la société Worldline au travers de sa filiale Santeos, concernant les données du dossier médical partagé (DMP) et la société Atos concernant toutes les autres données de *Mon espace santé*. Ces deux sous-traitants sont certifiés Hébergeurs de Données de Santé (HDS).

C'est l'intérêt public qui a motivé la création de cette plateforme : « *Mon espace santé* a pour objectif de promouvoir le rôle de chaque personne, tout au long de sa vie, dans la protection et l'amélioration de sa santé. Ce service public numérique domaine sécurisé vous permet de gérer vos données de santé en lien avec les acteurs des secteurs sanitaires, social et médico-social, favorisant ainsi la prévention, la coordination, la qualité et la continuité des soins. »[186]

On peut activer son espace santé ou s'opposer à sa création, mais pour cela il faut se connecter à son espace en utilisant le code d'accès envoyé par la CNAM. À la clôture d'un compte *Mon espace santé*, les données y seront conservées pendant dix ans.

On peut y déclarer ou non des professionnels de santé autorisés à consulter les documents, bien qu'en cas d'urgence une option « brise-glace » permette à des soignants d'y avoir accès. Une fois qu'une personne a créé son dossier sur *Mon espace santé*, tout professionnel de santé consulté par l'usager de la PDS, peut y déposer un document (compte-rendu de consultation, résultats d'analyse, nature des examens effectués, achat de médicaments etc). Il faut cependant remarquer que tous ne le font pas encore. L'usager

[184] Voir : https://www.casd.eu/.
[185] Voir : https://www.monespacesante.fr/
[186] Voir : https://www.monespacesante.fr/protection-donnees-personnelles

peut rendre visible ou masquer (sauf à l'auteur du document) aux professionnels de santé ses documents, son historique des soins » et ses directives anticipées. La confidentialité de chaque document peut être gérée, mais on ne peut pas supprimer un document déposé par un professionnel de santé. En cas de diagnostic et de document sensible l'accès est masqué au patient tant que celui-ci n'a pas eu d'entretien en face-à-face avec son médecin.

Mon Espace Santé n'est pas accessible pour les usages secondaires des données, en particulier pour la recherche.

Enfin *Mon espace santé* est présenté comme la composante française du futur « Espace européen des données de santé » qui fait actuellement l'objet d'une proposition de règlement du Parlement européen et du Conseil européen.[187]

7. INCa

L'Institut national du cancer (INCa) est un groupement d'intérêt public créé par la loi de santé publique du 9 août 2004, dans le cadre du Plan cancer 2003-2007, pour coordonner les actions de lutte contre le cancer. Il rassemble l'ensemble des acteurs de la lutte contre le cancer en France autour d'une double ambition : contribuer à diminuer la mortalité par cancer en France et améliorer la qualité de vie des personnes atteintes d'un cancer. L'État y est représenté par les ministères chargés de la santé et de la recherche. Parmi ses missions l'INCa recueille les informations les plus actualisées issues des différents producteurs de données, en réalise l'analyse et la synthèse pour produire une expertise pluridisciplinaire et partagée sur les questions relatives à la cancérologie : facteurs de risque, problématiques de démographie, radiothérapie, tests génétiques, évolution des organisations, impact des changements technologiques, accompagnement social…[188] La plateforme de données en cancérologie (PDC) développée par l'INCa est un entrepôt de données qui regroupe, dans les meilleures conditions de sécurité, des données issues de différentes sources. Unique en Europe par sa qualité, sa richesse et son volume, l'ambition est qu'elle devienne un outil indispensable de la recherche et du soin en cancérologie, pour soutenir et renforcer la prévention des cancers, l'amélioration des soins et de la qualité de vie des patients et la réduction des séquelles.[189]
Les relations de l'Institut national du cancer (INCa) avec les industries de santé sont strictement encadrées par des dispositions législatives, réglementaires et des règles internes approuvées par son comité de déontologie et d'éthique et son conseil d'administration. Ces règles visent à garantir l'indépendance de l'INCa lorsque, dans l'exercice de ses missions, l'institut collabore avec ces industries.[190]

8. Inserm– IReSP - Aviesan

L'Inserm (Institut national de la santé et de la recherche médicale) a une unité de service, le CépiDC (Centre d'épidémiologie sur les causes de décès) dont les missions sont la production des statistiques nationales sur les causes médicales de décès, la diffusion de ces statistiques, la conduite d'études et de recherches sur les données de mortalité. Ses données sont intégrées au SNDS et disponibles de manière ouverte sur le site du CépiDc[191]

[187] Voir : https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_fr
[188] Voir : https://solidarites-sante.gouv.fr/ministere/acteurs/agences-et-operateurs/article/inca-institut-national-du-cancer
[189] Voir : https://www.e-cancer.fr/Expertises-et-publications/La-plateforme-de-donnees-en-cancerologie
[190] Voir : https://www.e-cancer.fr/Institut-national-du-cancer/Deontologie-et-transparence-DPI/Le-cadre-de-la-deontologie
[191] Voir : https://www.cepidc.inserm.fr/

L'Inserm participe également avec l'IReSP (Institut de recherche en santé publique) et l'Aviesan (Alliance nationale pour les sciences de la vie et de la santé) au portail Epidémiologie - France qui propose un catalogue en ligne des principales bases de données en santé de source française qui peuvent être utiles au développement de la recherche et de l'expertise en santé publique[192].

Avec le projet France Cohortes[193] l'Inserm va mutualiser des moyens techniques et humains au service de onze de ses grandes cohortes, dont Constances (Annexe 4.9).

Il convient de souligner que l'Inserm a fait le choix d'utiliser la technologie Informatica pour sa PDS centralisée de données de santé

9. Constances

Constances[194] est une base de données et une PDS provenant de 220 000 participants volontaires français. Il s'agit donc une cohorte à grande échelle qui est ouverte à la communauté scientifique. Les équipes scientifiques françaises et internationales qui souhaitent bénéficier de Constances pour leurs propres recherches peuvent proposer des projets. Les équipes intéressées appartiennent essentiellement à des organismes publics de recherche comme l'Inserm, le CNRS, les universités. La possibilité de proposer des projets de recherche dans Constances est également ouverte à des équipes de recherche des entreprises industrielles, notamment du secteur de la santé. Tous les projets doivent avoir un objectif de santé publique et ceux qui sont susceptibles d'avoir un objectif de marketing sont exclus.

10. UK Biobank

UK Biobank[195] est une base de données et une PDS provenant d'un demi-million de participants volontaires britanniques. C'est donc une cohorte à grande échelle qui constitue une ressource de recherche unique en croisant des informations génétiques et des données sanitaires approfondies. La base de données est régulièrement enrichie de données supplémentaires et elle est accessible à l'échelle mondiale aux chercheurs agréés d'organismes publics ou privés qui entreprennent des recherches essentielles sur les maladies les plus courantes et les plus dangereuses.

11. Dawex

L'entreprise française Dawex[196] développe une plateforme d'échange de données, qui permet de distribuer et orchestrer un écosystème de données. Elle ne vend ni n'achète des données, mais met en relation des entreprises intéressées par la monétisation et la réutilisation des données. « La plateforme intègre un modèle de gouvernance des échanges de données regroupant le contrôle, la sécurité, la traçabilité, l'octroi de licence et la conformité réglementaire. Elle peut être opérée en mode centralisé, distribué ou décentralisé et propose plusieurs modèles économiques à ses participants : gratuit, par abonnement, par transaction[197]». Dawex propose un certain nombre d'outils pour aider les fournisseurs et les utilisateurs de données à bien évaluer la nature des données échangées. Elle propose également des outils d'échantillonnage pour lutter contre les biais

[192] Voir : https://epidemiologie-france.aviesan.fr/
[193] Voir : https://www.inserm.fr/actualite/france-cohortes-comment-perenniser-outil-recherche-exceptionnel/
[194] Voir : https://www.constances.fr/.
[195] Voir : https://www.ukbiobank.ac.uk/.
[196] Voir : https://www.dawex.com/.
[197]Voir : https://www.dawex.com/fr/data-exchange-platform/.

de représentativité des données. Actuellement Dawex n'est pas spécialisé en données de santé mais est déjà utilisé en agriculture : API-AGRO est un centre de partage de données agricoles qui repose sur la technologie Dawex.

12. Salus-Co-op

Salus-Co-op[198] est une coopérative citoyenne espagnole, créée en 2017, qui propose à ses membres de donner accès à leurs informations médicales pour des projets de recherche en santé portés par des institutions sans but commercial, pourvu qu'elles partagent leurs données de recherche librement et gratuitement, et sous réserve que les donateurs ne retirent pas leur accord. Avec l'application Salus-Co-op, les données sont pseudonymisées, cryptées de bout en bout et transitent directement entre les utilisateurs et les chercheurs des projets auxquels ils participent[199].

13. Healthbank

En Suisse, la plateforme d'échanges de données de santé de Healthbank[200] est une initiative coopérative qui propose à chaque personne de partager de ses données de santé avec qui elle veut de façon anonyme et sécurisée, éventuellement en monétisant cet accès. Les utilisateurs restent propriétaires de leurs données personnelles de santé et peuvent décider à tout moment et quelle qu'en soit la raison d'interrompre le partage de leurs données[201]. Dans la version de base, l'ouverture d'un compte est gratuite, mais on peut devenir membre de la coopérative en achetant une part (100 CHF). La plateforme sert d'intermédiaire (payant) entre les chercheurs et les membres de la plateforme. C'est elle qui anonymise les données, se fait payer par les chercheurs et paie les membres qui vendent l'accès à leurs données.

14. Doctolib

Doctolib[202] est une entreprise française qui a commencé en 2013 à distribuer en France, puis en Italie et en Allemagne, une application de gestion des rendez-vous réservée aux professionnels de la santé ainsi qu'un service de prise de rendez-vous en ligne, destiné aux patients. Progressivement Doctolib a développé un service de gestion de la patientèle et de ses données pour les médecins. Mais en 2021, plusieurs syndicats de médecins ont déposé devant le Conseil d'État un recours en raison de l'utilisation par Doctolib des services d'hébergement d'Amazon Web Services (AWS). Cependant le Conseil d'État a répondu en affirmant que « Doctolib a [...] mis en place un dispositif de sécurisation des données hébergées par la société AWS Sarl reposant sur un tiers de confiance situé en France afin d'empêcher la lecture des données par des tiers » et a validé ainsi la conformité du système d'hébergement mis en place par Doctolib.[203]

[198] Voir : https://www.saluscoop.org/

[199] *La Croix*, (Débat), *Faut-il partager ses données au nom de l'intérêt général ? Des risques de pratique discriminatoires,* 28 juin 2022.

[200] Voir : https://www.healthbank.coop/

[201] Voir : https://www.healthbank.coop/#how-it-works.

[202] Voir : https://www.doctolib.fr/.

[203] Conseil d'Etat, décision de justice du 12 mars 2021 [https://www.conseil-etat.fr/actualites/le-juge-des-referes-ne-suspend-pas-le-partenariat-entre-le-ministere-de-la-sante-et-doctolib-pour-la-gestion-des-rendez-vous-de-vaccination-contre].

Annexe 5 : Données issues de la recherche médicale : quel cadre légal ?

Les formalités et démarches concernant les données issues de recherches impliquant la personne humaine dépendent du contexte de la recherche. La CNIL distingue en effet deux périmètres différents : la recherche en interne (effectuée auprès de patients dans le cadre de leur suivi thérapeutique et médical par l'équipe de soins et pour son usage exclusif) et la recherche multicentrique ou impliquant que les données soient rendues accessibles à des tiers (par exemple, pour une thèse ou un mémoire).

Dans le premier cas, les recherches sont encadrées par les dispositions du RGPD et doivent répondre aux exigences du chapitre IX de la loi Informatiques et Libertés. Dans le second cas, la procédure nécessite davantage de points de vigilance afin que la publication de la recherche ne puisse pas permettre d'identifier directement ou indirectement les personnes concernées. Cette démarche insiste sur la responsabilisation des acteurs et exige une analyse d'impact sur la protection des données (AIPD) dès lors que leur utilisation peut entraîner un risque élevé pour les droits et libertés des personnes concernées.

Actuellement, en France, la recherche clinique est régie par la loi Jardé, qui encadre les « recherches impliquant la personne humaine (RIPH) » portant sur un dispositif médical, un médicament ou un autre produit de santé. Ces recherches sont classifiées en trois catégories en fonction du risque encouru par le participant :

- les RIPH 1 (officiellement dénommées Recherches Interventionnelles, RI), concernent les essais médicamenteux ou de nouveaux dispositifs implantables et qui passent par une autorisation spécifique auprès de L'Agence nationale de sécurité du médicament et des produits de santé (ANSM) ; ces recherches mettent en œuvre une intervention non habituellement pratiquée sur la personne humaine ;
- les RIPH 2 (officiellement Recherche Interventionnelle à Risques et Contraintes Minimes, RIRCM) comportent une intervention sur la personne, dont la liste est fixée par arrêté du ministre de la Santé ;
- les RIPH 3 (officiellement Recherche Non Interventionnelle, RNI) ne comportent aucun risque, les actes étant pratiqués de manière habituelle, même si un arrêté du ministre de la Santé définit les actes autorisés pour cette catégorie de recherche, ce qui peut amener une certaine confusion.

Les Comités de protection des personnes (CPP) coordonnés par la Commission nationale des recherches impliquant la personne humaine (CNRIPH)[204], sont chargés « d'émettre un avis préalable sur les conditions de validité de toute recherche impliquant la personne humaine (essai ou expérimentation), au regard des critères définis par l'article L 1123-7 du Code de la Santé Publique (CSP). Les comités s'assurent notamment que la protection des participants à la recherche impliquant la personne humaine est assurée (information préalable, recueil du consentement, période d'exclusion, délai de réflexion…), que la recherche est pertinente, que l'évaluation du rapport bénéfice/risque est satisfaisant »[205].

Néanmoins il existe de nombreux domaines de recherche utilisant des données de santé qui ne relèvent pas des RIPH : ce sont les recherches qui nécessitent la réutilisation des données de santé à caractère personnel issues notamment des dossiers médicaux, du

[204] Voir : https://solidarites-sante.gouv.fr/systeme-de-sante-et-medico-social/innovation-et-recherche/article/la-commission-nationale-des-recherches-impliquant-la-personne-humaine-cnriph.
[205] Voir : https://www.iledefrance.ars.sante.fr/comites-de-protection-des-personnes-cpp.

SNDS ou de cohortes. Ces données peuvent être réutilisées si le patient en est informé et ne s'y oppose pas.

Annexe 6 : Auditions

Naomi Allen, responsable scientifique de UK Biobank ;
Régis Aubry, médecin en soins palliatifs, membre du CCNE ;
Emmanuel Bacry, directeur scientifique du GIP-PDS (*Health Data Hub*);
Sarah Benichou, cheffe de pôle au département promotion de l'égalité et de l'accès au droit (Défenseur des droits) ;
Eric Bothorel, député de Côtes d'Armor (Renaissance), rapporteur de la mission parlementaire « Pour une politique publique de la donnée » ;
Erik Boucher de Crèvecoeur, ingénieur expert à la Commission nationale de l'informatique et des libertés (CNIL) ;
Éric Chenut, président de la Mutualité française ;
Marie Citrini, représentante des usagers à l'Assistance Publique des Hôpitaux de Paris (AP-HP) ;
Stéphanie Combes, directrice du GIP-PDS ;
Caroline Cormet-Fraigneau, vice-présidente en charge du développement d'OVHcloud ;
Marc Cuggia, médecin de santé publique, professeur des universités - praticien hospitalier (PU-PH) en informatique médicale et co-pilote de la mission de préfiguration du GIP-PDS ;
Annabelle Cumyn, membre du groupe de recherche interdisciplinaire en informatique de la santé, Université de Sherbrooke et présidente du Comité d'éthique de la recherche du CIUSSS de l'Estrie ;
Arthur Dauphin, chargé de mission à France Assos Santé ;
Jean-François Ethier, directeur du Centre interdisciplinaire de recherche en informatique de la santé de l'Université de Sherbrooke ;
Valérie Fontaine, chargée des partenariats auprès du Défenseur des droits ;
Guy Fournier, directeur Secteur Public et collectivités d'OVHcloud ;
Jérémy Greene, professeur d'histoire de la médecine à la Johns Hopkins University ;
Caroline Guillot, adjointe à la direction citoyenne du GIP-PDS ;
Hélène Guimiot-Bréaud, cheffe du Service de la santé à la Commission nationale de l'informatique et des libertés (CNIL) ;
Anne Gysenbergh-Houal, responsable des collaborations de recherche académique et industrielle, délégation à la recherche clinique et à l'innovation de l'AP-HP ;
Claudine Jacob, directrice de la protection des droits-affaires judiciaires auprès du Défenseur des droits ;
Nicolas Kanhonou, directeur de la promotion de l'égalité et de l'accès aux droits auprès du Défenseur des droits ;
Benoît Labarthe, responsable du département Partenariats et Innovations à la Direction de la Recherche et de l'Innovation, pôle affaires médicales, recherche et stratégie territoriale au CHU de Nantes ;
Laurent Lafaye, co-fondateur de la société Dawex ;
Philippe Latombe, député de Vendée (MoDem) et rapporteur de la mission parlementaire « Bâtir et protéger une souveraineté numérique nationale et européenne » ;
Karine Lefeuvre, professeure de droit des personnes vulnérables à l'EHESP et vice-présidente du CCNE ;
Franck Lethimonnier, directeur du Consortium de valorisation thématique de l'Alliance Aviesan ;
Laura Létourneau, déléguée ministérielle au Numérique de Santé (DNS) ;
Pierre Lombrail, professeur des Universités-Praticien Hospitalier en santé publique à l'Université Paris 13, rapporteur du GT « données massives en santé » du Comité d'Éthique de l'Inserm (CEI) ;

Jacques Lucas, président de l'Agence numérique de santé (ANS)
Emmanuel Meyrieux, responsable sécurité clients d'OVHcloud ;
Catherine Morin-Desailly, sénatrice de la Seine Maritime (Union centriste, groupe UCI-UC) investie sur la question de la gouvernance européenne du numérique ;
Frédéric Ossant, chef de projet de la plateforme Ouest Data Hub ;
Adrien Parrot, président de InterHop ;
Denis Paul, chef de projets à OVHcloud ;
Valérie Peugeot, chercheuse au sein d'Orange Labs et Présidente de l'association Vecam, commissaire en charge des données de santé à la CNIL ;
Christelle Rebillet, responsable de Pôle, comité français d'accréditation ;
Guillaume Ruty, directeur informatique d'OVHcloud ;
Brigitte Seroussi, directrice de projets à la Délégation ministérielle au Numérique de Santé (DNS) ;
Catherine Simonin, membre du bureau de France Assos Santé, membre de la Ligue nationale contre le cancer ;
Hubert Tardieu, président du conseil d'administration de Gaia-X ;
Fabrice Tocco, co-fondateur de la société Dawex ;
Celia Zolinsky, professeur de droit à l'université Panthéon -Sorbonne, membre du CNPEN

Remerciements :
Bastien Rance (Imagine)